1. Introduction
2. Preview of Comm System Components
3. Preliminaries
4. The Communications Channel
5. Analog Communication
6. Random Signals and Noise
7. Complex-Baseband Equivalent Channel
8. Digital Communication
9. Discrete-Time Implementation of Digital Communication
10. Error Analysis of Digital Communications

Introduction

This module presents introductory material on data communication, focusing on a high-level summary/comparison of analog and digital communication.

Goal: Transmit a message from one location to another.

When message is a...
continuous waveform, use analog comm (e.g., FM radio),
sequence of numbers, use digital comm (e.g., mp3 file),
but keep in mind that the sequence of numbers might represent a sampling of a continuous waveform (as in the case of digital audio).

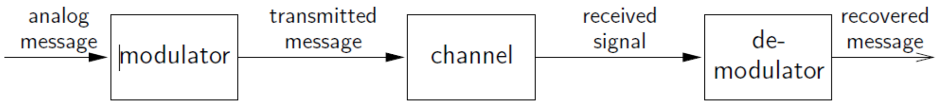
Typical communication media:

twisted pair wire	(e.g., telephone$_A$)
coaxial cable	(e.g., TV$_{A,D}$, data$_D$)
fiber optic cable	(e.g., ethernet$_D$)
EM waves	(e.g., cellular phones$_{A,D}$, WiFi$_D$, TV$_{A,D}$)
water waves	(e.g., underwater network$_{A,D}$)
power lines$_{A,D}$	
compact disc$_D$	
hard drive$_D$	
magnetic tape$_{A,D}$	

where $A = \text{analog}$ and $D = \text{digital}$.

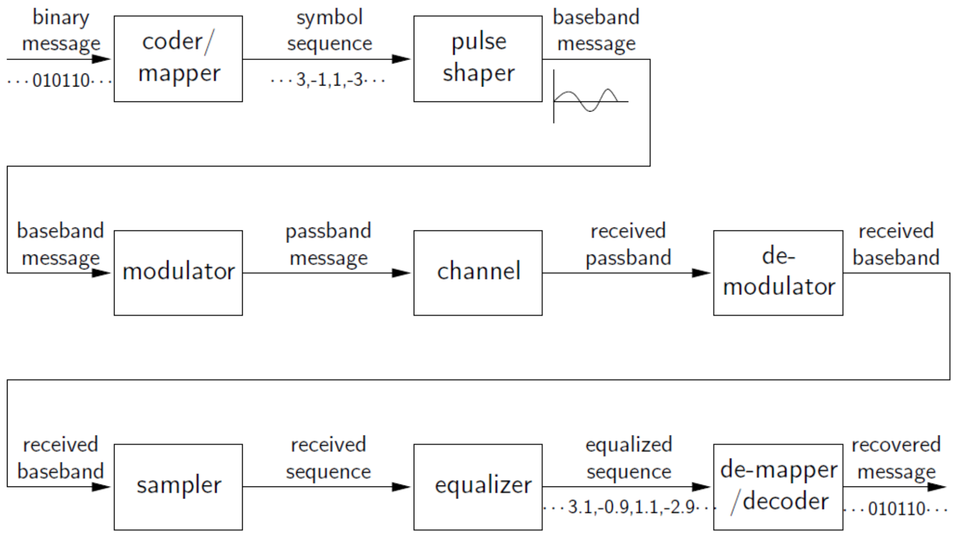
Note that, whether the message signal is discrete-time or continuous-time, the transmitted signal is continuous-time!

Analog Communication



- Perfect recovery is impossible in the presence of noise!

Digital Communication



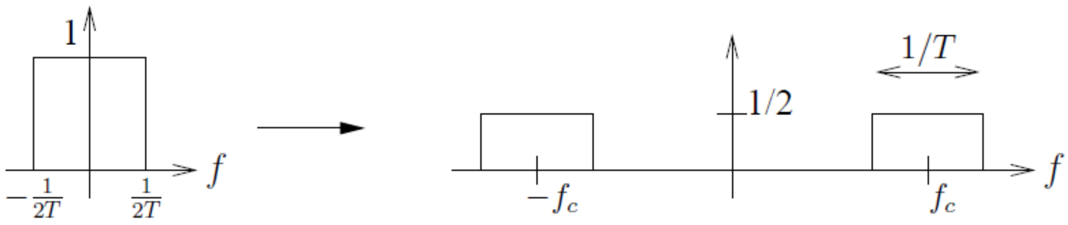
- A digital message is converted to an analog message coding and pulse-shaping, and then transmitted using analog modulation. To recover the message, the received signal is demodulated, sampled, and digitally processed.
- Perfect recovery is possible even in the presence of noise!

Preview of Comm System Components

This module gives a brief description of important communication system components: the modulator, the coder/mapper, and the pulse shaper.

The Modulator:

- Translates “baseband” analog message to “passband”:



  where f_c is the “carrier frequency.”
- There are two principal motivations for doing this:

  1. Often we want to communicate several messages simultaneously (e.g., TV, radio, voice). It's difficult or impossible to do this if they overlap in frequency!
  2. Wireless EM transmission/reception is much easier at higher frequencies, since need antenna length $> \frac{\lambda}{10}$. ($\lambda = \frac{c}{f_c}$ is wavelength and c=3e8 m/s speed of light.)

system	transmission band	$\lambda/10$
VHF (TV)	30–300 MHz	1–0.1 m
UHF (TV)	0.3–3 GHz	10–1 cm
cellular	824–960 MHz	3 cm
WiFi	2.4 GHz	1 cm

> Notice that practical antenna length determines where different signal types can be transmitted.

The Coder/Mapper:

- Coder transforms sequence of message bits into an error-resiliant sequence of coded bits.
- Mapper transforms coded bits into discrete “symbols.” Ex: If the “symbol alphabet” is $\{-3, -1, 1, 3\}$ and the symbol mapping is

bits	symbol
00	3
01	-1
10	1
11	-3

then ASCII text would be transmitted via

letter	ASCII code				symbol sequence			
a	01	10	00	01	-1	1	-3	-1
b	01	10	00	10	-1	1	-3	1

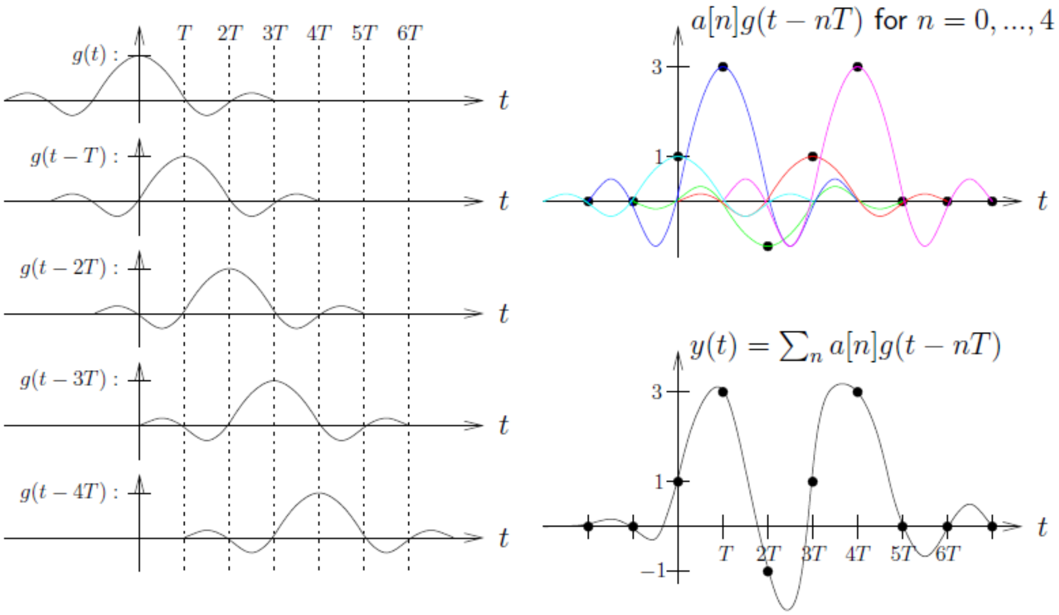
c	01	10	00	11	-1	1	-3	3
d	01	10	01	00	-1	1	-1	-3
⋮	⋮				⋮			

Pulse Shaper:

- Converts symbol sequence into a continuous waveform.
- In linear modulation schemes, the time-*n* symbol $a[n]$ scales a nT-delayed version of pulse $g(t)$:
  Equation:

$$\begin{aligned} m(t) &= \sum_n a[n] g(t-nT) \text{”baseband message”} \\ T &= \text{”symbol period”} \end{aligned}$$

Ex: Say symbol sequence is $[1, 3, -1, 1, 3]$. Then



Preliminaries
The module provides a review of the background concepts needed for the study of analog and digital communications: the Fourier transform, various definitions of bandwith, the Dirac delta, frequency-domain representation of sinusoids, frequency-domain plotting in Matlab, linear time invariant (LTI) systems, linear filtering, lowpass filters, and Matlab design of lowpass filters.

Fourier Transform (FT)

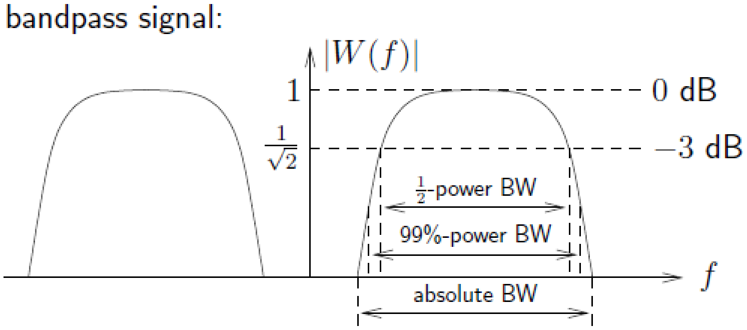
Definition:
Equation:

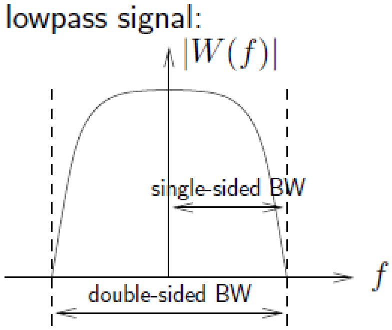
$$\begin{aligned} W(f) &= \int_{-\infty}^{\infty} w\,(t)e^{-j2\pi ft}dt = \mathscr{F}\,\{w\,(t)\} \\ w(t) &= \int_{-\infty}^{\infty} W\,(f)e^{j2\pi ft}df = \mathscr{F}^{-1}\,\{W\,(f)\}. \end{aligned}$$

Properties:

- Linearity: $\mathscr{F}\,\{c_1 w_1\,(t) + c_2 w_2\,(t)\} = c_1 W_1\,(f) + c_2 W_2\,(f)$.
- Real-valued $w(t)\ \Rightarrow \begin{cases} \text{conjugate symmetric} W(f) \\ |W(f)| \text{symmetric around} f = 0 \end{cases}.$

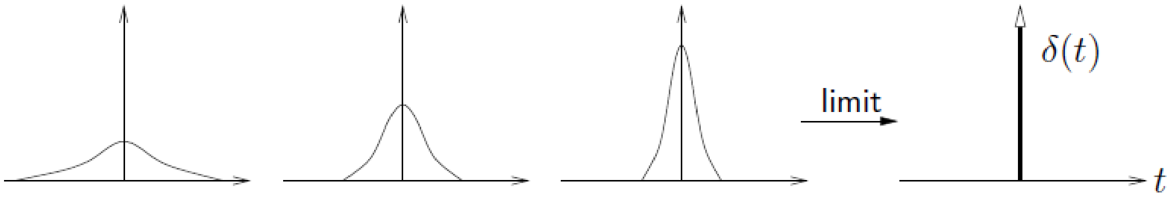
“Bandwidth”





Dirac Delta (or “continuous impulse”) $\delta(\cdot)$

An infinitely tall and thin waveform *with unit area*:



that's often used to “kick” a system and see how it responds.
Key properties:

1. Sifting: $\int_{-\infty}^{\infty} w(t)\delta(t-q)dt = w(q)$.
2. Time-domain impulse $\delta(t)$ has a flat spectrum:
   Equation:

$$\mathscr{F}\{\delta(t)\} = \int_{-\infty}^{\infty} \delta(t)e^{-j2\pi ft}dt \;=\; 1 \text{ (for all } f\text{)}.$$

3. Freq-domain impulse $\delta(f)$ corresponds to a DC waveform:
   Equation:

$$\mathscr{F}^{-1}\{\delta(f)\} = \int_{-\infty}^{\infty} \delta(f) e^{j2\pi ft} df \ = \ 1 \text{ (for all } t\text{)}.$$

Frequency-Domain Representation of Sinusoids

Notice from the sifting property that

Equation:

$$\mathscr{F}^{-1}\{\delta(f-f_o)\} = \int_{-\infty}^{\infty} \delta(f-f_o) e^{j2\pi ft} df = e^{j2\pi f_o t}.$$

Thus, Euler's equations

Equation:

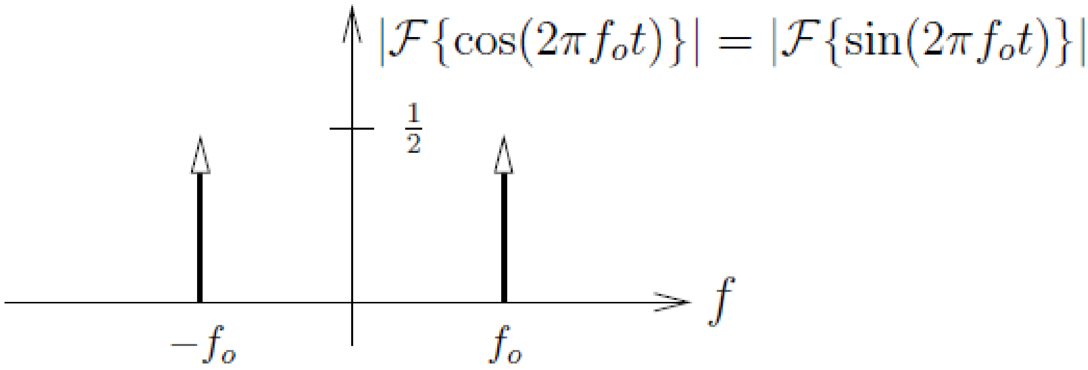
$$\begin{aligned} \cos(2\pi f_o t) &= \tfrac{1}{2} e^{j2\pi f_o t} + \tfrac{1}{2} e^{-j2\pi f_o t} \\ \sin(2\pi f_o t) &= \tfrac{1}{2j} e^{j2\pi f_o t} - \tfrac{1}{2j} e^{-j2\pi f_o t} \end{aligned}$$

and the Fourier transform pair $e^{j2\pi f_o t} \leftrightarrow \delta(f-f_o)$ imply that

Equation:

$$\begin{aligned} \mathscr{F}\{\cos(2\pi f_o t)\} &= \tfrac{1}{2}\delta(f-f_o) + \tfrac{1}{2}\delta(f+f_o) \\ \mathscr{F}\{\sin(2\pi f_o t)\} &= \tfrac{1}{2j}\delta(f-f_o) - \tfrac{1}{2j}\delta(f+f_o). \end{aligned}$$

Often we draw this as



$|\mathcal{F}\{\cos(2\pi f_o t)\}| = |\mathcal{F}\{\sin(2\pi f_o t)\}|$
$\frac{1}{2}$
f
$-f_o$
f_o

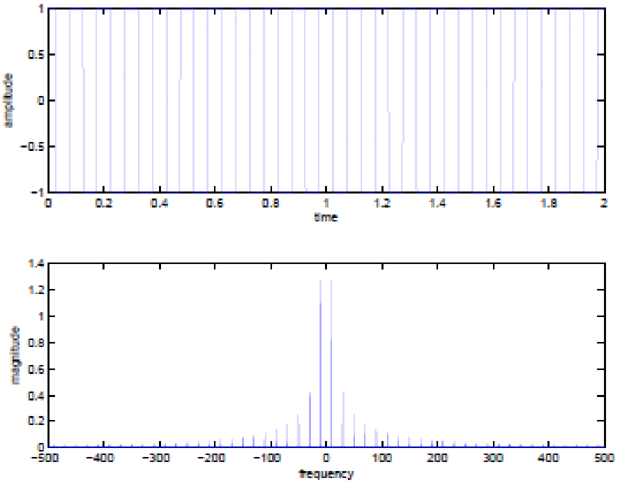
Frequency Domain via MATLAB

Fourier transform requires evaluation of an integral. What do we do if we can't define/solve the integral?

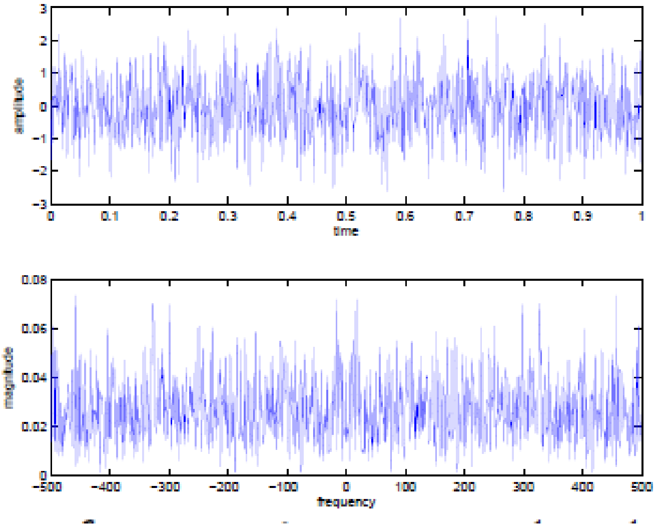
1. Generate (rate-$\frac{1}{T_s}$) sampled signal in MATLAB.
2. Plot magnitude of Discrete Fourier Transform (DFT) using `plottf.m` (from course webpage).

Square-wave example:

```
f = 10;
t_max = 2;
Ts = 1/1000;
t = 0:Ts:t_max;
x =
sign(cos(2*pi*f*t));
plottf(x,Ts);
```



Noise-wave example

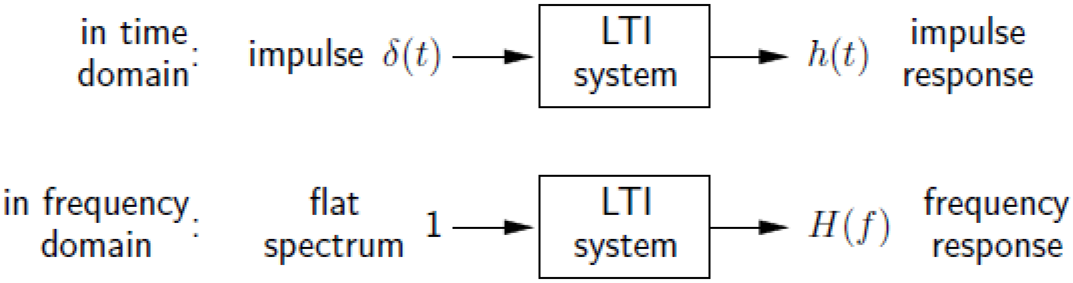


```
t_max = 1;
Ts = 1/1000;
x =
randn(1,t_max/Ts);
plottf(x,Ts);
```

Notice that `plottf.m` only plots frequencies $f \in \left[-\frac{1}{2T_s}, \frac{1}{2T_s}\right)$.

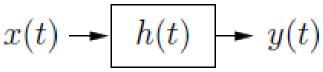
Linear Time-Invariant (LTI) Systems

An LTI system can be described by either its “impulse response” $h(t)$ or its “frequency response” $H(f) = \mathscr{F}\{h(t)\}$.



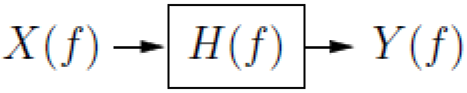
Input/output relationships:

- Time-domain: Convolution with impulse response $h(t)$



$$y(t) = h(t) * x(t) = \int_{-\infty}^{\infty} h(t-\tau)x(\tau)d\tau$$

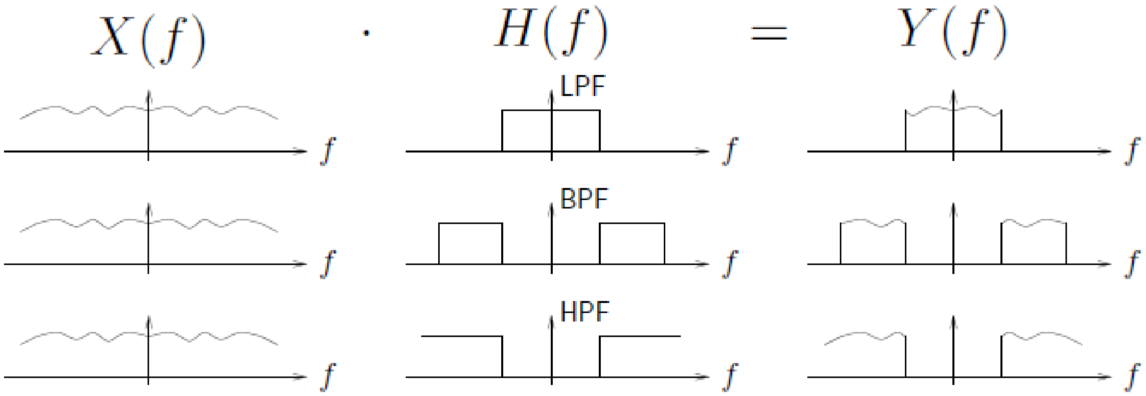
- Freq-domain: Multiplication with freq response $H(f)$



$$Y(f) = H(f)X(f)$$

Linear Filtering

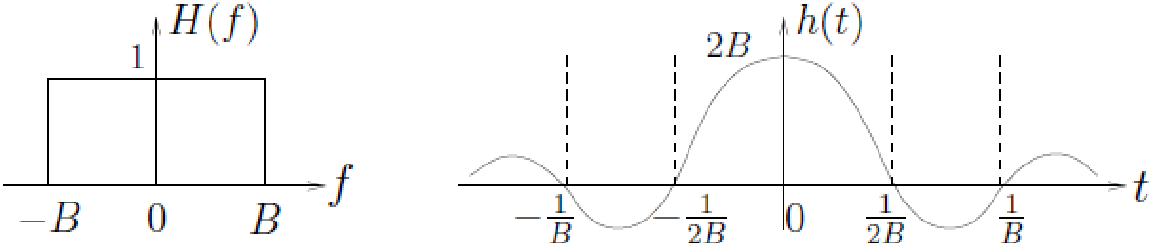
Freq-domain illustration of LPF, BPF, and HPF:



Lowpass Filters

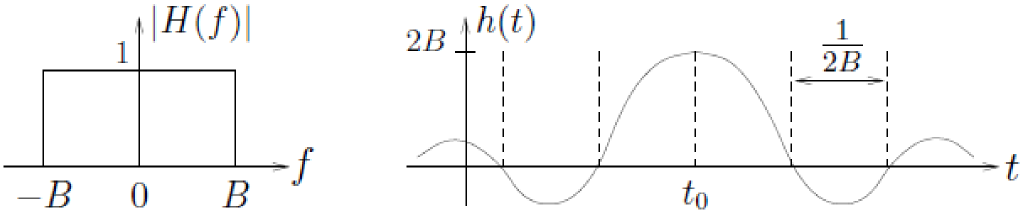
Ideal non-causal LPF (using sinc $(x) := \frac{\sin(\pi x)}{\pi x}$):

$$H(f) = \begin{cases} 1 & |f| \leq B \\ 0 & |f| > B \end{cases} \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad h(t) = 2B\,\text{sinc}(2Bt)$$

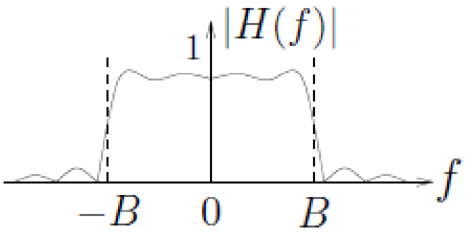


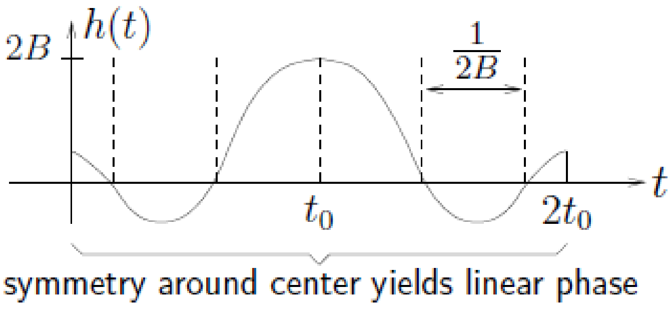
Ideal LPF with group-delay t_o:

$$H(f) = \begin{cases} e^{-j2\pi f t_0} & |f| \leq B \\ 0 & |f| > B \end{cases} \quad \overset{\mathcal{F}}{\longleftrightarrow} \quad h(t) = 2B\,\text{sinc}\left(2B(t-t_0)\right)$$



A causal linear-phase LPF with group-delay t_o:



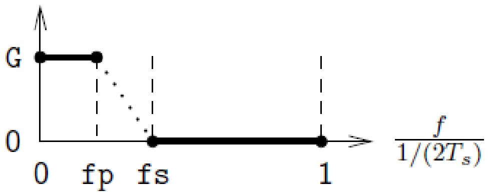


but MATLAB can give better causal linear-phase LPFs...

In MATLAB, generate $\frac{1}{T_s}$-sampled LPF impulse response via

`h = firls(Lf, [0,fp,fs,1], [G,G,0,0])/Ts;`

where...



$\texttt{Lf}+1$ = impulse response length
$\{0, \texttt{fp}\}, \{\texttt{fs}, 1\}$ = *normalized* freq pairs
$\{\texttt{G}, \texttt{G}\}, \{0, 0\}$ = corresp. magnitude pairs

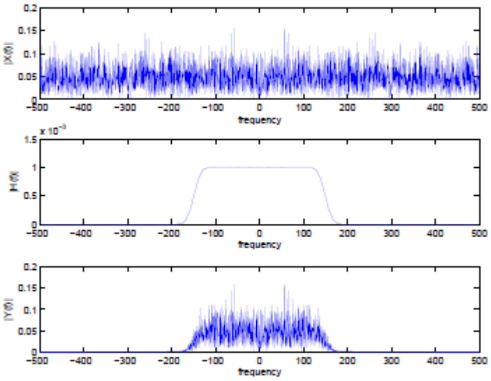
The commands `firpm` and `fir2` have the same interface, but yield slightly different results (often worse for our apps).

In MATLAB, perform filtering on $\frac{1}{T_s}$-sampled signal `x` via

`y = Ts*filter(h,1,x);` or `y = Ts*conv(h,x);`

```
t_max = 3; Ts = 1/1000;
x = randn(1,t_max/Ts);
```

```
h = firls(100,[0,0.2,0.4,1],
[1,1,0,0])/Ts;
y = Ts*filter(h,1,x);
subplot(3,1,1);
plottf(x,Ts,’f’);
ylabel(’|X(f)|’)
subplot(3,1,2);
plottf(h,Ts,’f’);
ylabel(’|H(f)|’)
subplot(3,1,3);
plottf(y,Ts,’f’);
ylabel(’|Y(f)|’)
```

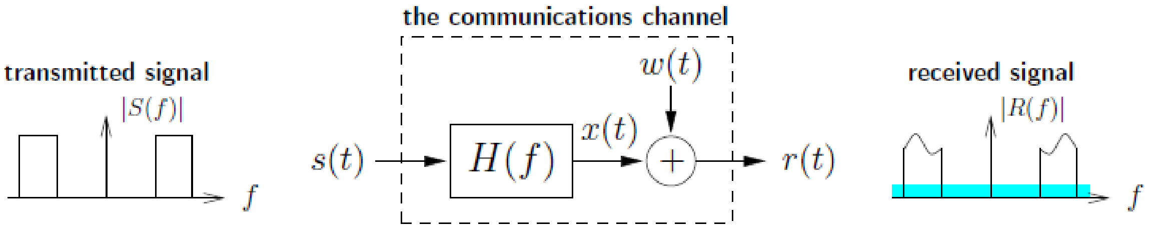


Note: The routines `firls,firpm,fir2` generate *causal* linear-phase filters with group delay $= \frac{}{2}$ samples. Thus, the filtered output `y` will be delayed by $\frac{}{2}$ samples relative to `x` .

The Communications Channel

This module gives a brief introduction to the communications channel, which is modeled here as a linear filtering plus an additive noise.

The effects of signal propagation are usually modeled as:



where

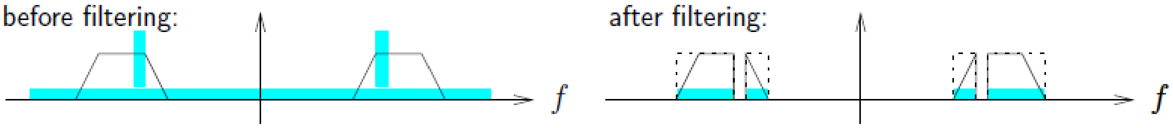
$H(f)$: linear filtering due to multipath propagation
$w(t)$: additive noise/interference.

Noise/interference sources include:

- electronic circuitry ("thermal/shot noise" or "quantization noise"); usually broadband,
- other comm systems ("multi-access interference" or "co-channel interference"); broadband or narrowband.

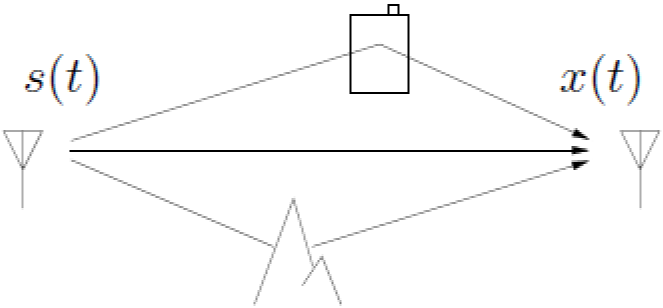
SNR can be improved with appropriate filtering at receiver:

- signal-to-noise ratio (SNR) $= \frac{\text{signal energy}}{\text{noise energy}} = \frac{\mathscr{E}_s}{\mathscr{E}_n}$
- roughly speaking, average SNR can be improved by filtering out the frequencies dominated by noise:



Filtering due to Multipath Propagation

The signal may propagate along paths with different lengths:



Since different path lengths imply different path delays:

Equation:

$$x\left(t\right) = c_1 s\left(t - \tau_1\right) + c_2 s\left(t - \tau_2\right) + \cdots + c_N s\left(t - \tau_N\right),$$

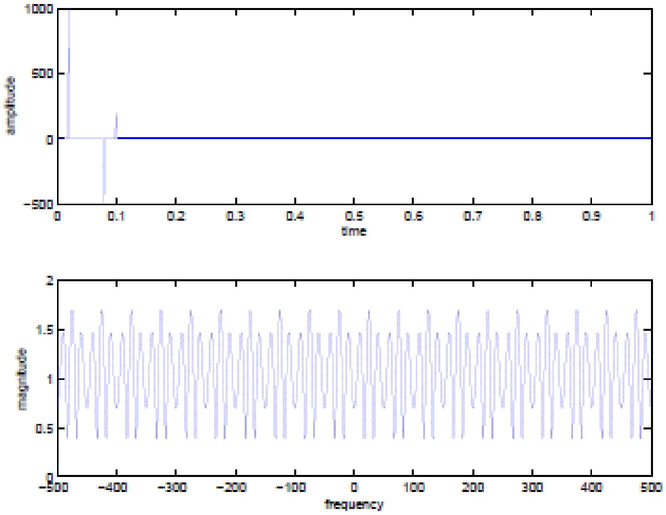
which can be written as $x(t) = s(t) * h(t)$ for

Equation:

$$h\left(\tau\right) = c_1 \delta\left(\tau - \tau_1\right) + c_2 \delta\left(\tau - \tau_2\right) + \cdots + c_N \delta\left(\tau - \tau_N\right).$$

The result is an $|H(f)|$ that varies with *f*, implying frequency-dependent signal attenuation:

```
t_max = 1;
Ts = 1/1000;
c1 = 1; tau1 = 0.02;
c2 = -0.5; tau2 =
0.08;
c3 = 0.2; tau3 =
0.10;
h =
zeros(1,t_max/Ts);
h(tau1/Ts) = c1/Ts;
h(tau2/Ts) = c2/Ts;
h(tau3/Ts) = c3/Ts;
plottf(h,Ts);
```



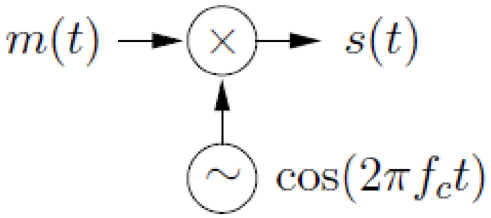
Analog Communication

This module describes basic analog modulation techniques, including amplitude modulation (AM) with suppressed carrier, AM with a pilot tone or carrier tone, quadrature AM (QAM), vestigial sideband modulation (VSB), and frequency modulation (FM). Various demodulation techniques are also discussed, including envelope detection and the discriminator. Application examples include NTSC television and FM radio (both mono and stereo).

1. Amplitude modulation (AM)
2. Quadrature amplitude modulation (QAM)
3. Vestigial sideband modulation (VSB)
4. Frequency modulation (FM)

AM with “suppressed carrier”

AM of real-valued message $m(t)$ (e.g., music) is



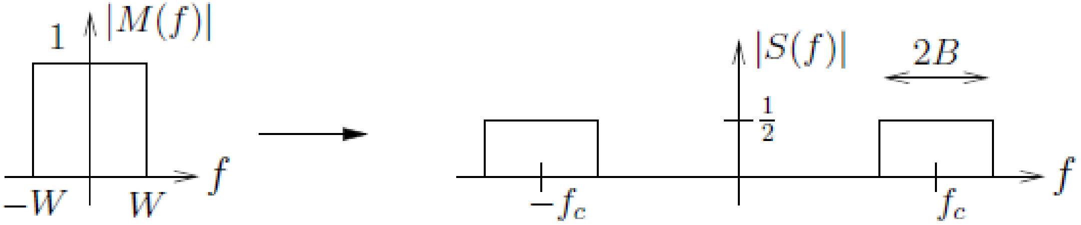
$$s(t) = m(t)\cos(2\pi f_c t),$$

$$f_c = \text{carrier freq.}$$

Euler's $\cos\left(2\pi f_c t\right) = \frac{1}{2}\left[e^{j2\pi f_c t} + e^{-j2\pi f_c t}\right]$ then implies

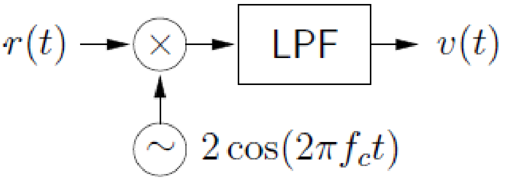
Equation:

$$\begin{aligned} S\left(f\right) &= \int_{-\infty}^{\infty} m\left(t\right)\,\cos\left(2\pi f_c t\right)e^{-j2\pi ft}dt \\ &= \frac{1}{2}\int_{-\infty}^{\infty} m\left(t\right)e^{-j2\pi(f-f_c)t}dt + \frac{1}{2}\int_{-\infty}^{\infty} m\left(t\right)e^{-j2\pi(f+f_c)t}dt \\ &= \frac{1}{2}M\left(f-f_c\right) + \frac{1}{2}M\left(f+f_c\right). \end{aligned}$$



Because $m(t) \in \mathbb{R}$, know $|M(f)|$ symmetric around $f = 0$, implying the AM transmitted spectrum below f_c is redundant! This motivates the QAM and VSB modulation schemes...

With f_c known, AM demodulation can be accomplished by:



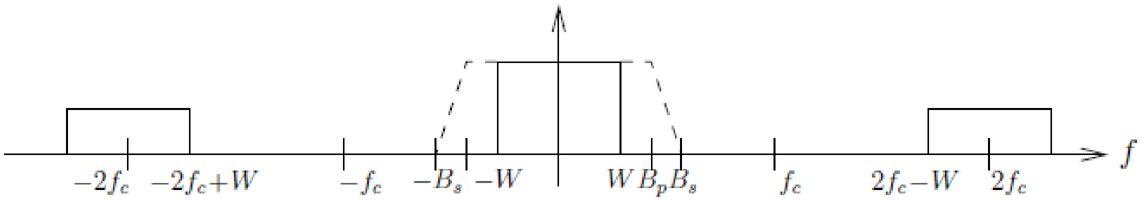
$$v(t) = \text{LPF}\{r(t) \cdot 2\cos(2\pi f_c t)\}.$$

For a trivial noiseless channel, we have $r(t) = s(t)$, so that

Equation:

$$\begin{aligned} v(t) &= \text{LPF}\{s(t) \cdot 2\cos(2\pi f_c t)\} \\ &= \text{LPF} \quad m(t) \cdot \underbrace{2\cos^2(2\pi f_c t)}_{1+\cos(2\pi \cdot 2 f_c t)} \\ &= \text{LPF}\{m(t) + m(t)\cos(2\pi \cdot 2 f_c t)\} \\ &= m(t), \end{aligned}$$

assuming a LPF with passband cutoff $B_p \geq W$ Hz and stopband cutoff $B_s \leq 2f_c - W$ Hz:



Note that we've assumed perfectly synchronized oscillators!

When the receiver oscillator has {freq,phase} offset $\{\gamma, \varphi\}$:
Equation:

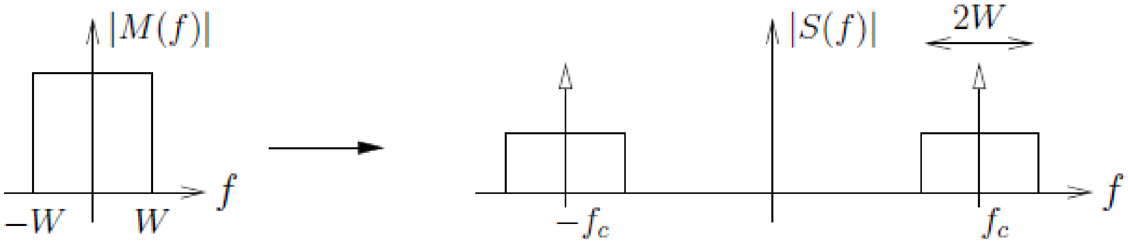
$$
\begin{aligned}
v(t) &= \text{LPF}\quad m(t)\cos(2\pi f_c t)\cdot 2\cos(2\pi(f_c+\gamma)t+\varphi)\\
&\qquad\qquad \cos(2\pi\gamma t+\varphi)+\cos(2\pi(2f_c+\gamma)t+\varphi)\\
&= m(t)\cos(2\pi\gamma t+\varphi)\quad .\\
&\qquad \text{time-varying attenuation!}
\end{aligned}
$$

Note:a freq offset of $\lambda = \frac{\nu f_c}{c}$ Hz can occur when there is relative velocity of *v* m/s between transmitter and receiver.

AM with “pilot tone” or “carrier tone”

It's common to include a pilot/carrier tone with frequency| f_c:
Equation:

$$
\begin{aligned}
s(t) &= m(t)\cos(2\pi f_c t) + A\cos(2\pi f_c t)\\
&\qquad\qquad\qquad \text{pilot/carrier tone}\\
&= [m(t)+A]\cos(2\pi f_c t)\\
S(f) &= \tfrac{1}{2}[M(f-f_c)+M(f+f_c)+A\delta(f-f_c)+A\delta(f+f_c)]
\end{aligned}
$$



Advantage: aids receiver with carrier synchronization.

Disadvantage: consumes transmission power.

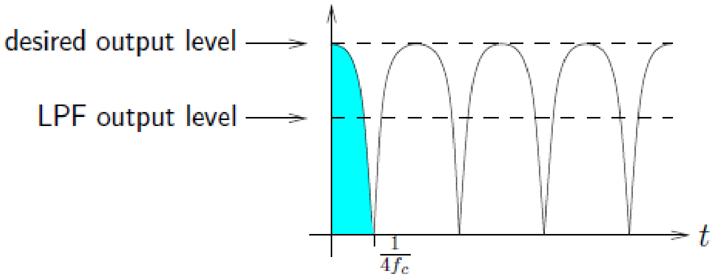
While modern systems choose $A \ll \max |m(t)|$, many older systems use $A > \max |m(t)|$, known as “large carrier AM,” allowing reception based on *envelope detection*:

Equation:

$$\begin{aligned} v(t) &= \frac{\pi}{2}\text{LPF}\{|r(t)|\} - A \\ &\approx m(t) \quad \text{(with a trivial channel)} \end{aligned}$$

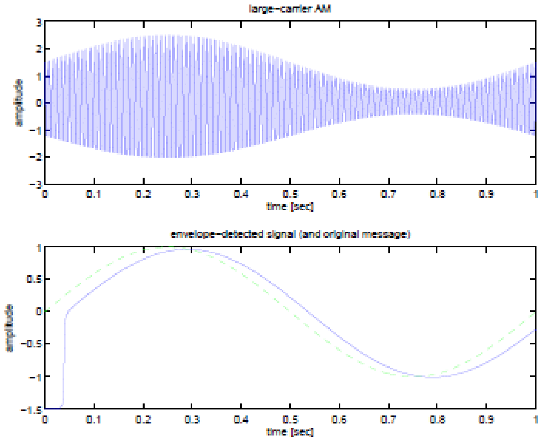
where $|\cdot|$ can be easily implemented using a diode.

The gain $\frac{\pi}{2}$ above makes up for the loss incurred when LPFing the rectified signal:



$$\frac{\int_0^{\frac{1}{4f_c}} \cos(2\pi f_c t)dt}{\frac{1}{4f_c}} = \frac{2}{\pi}.$$

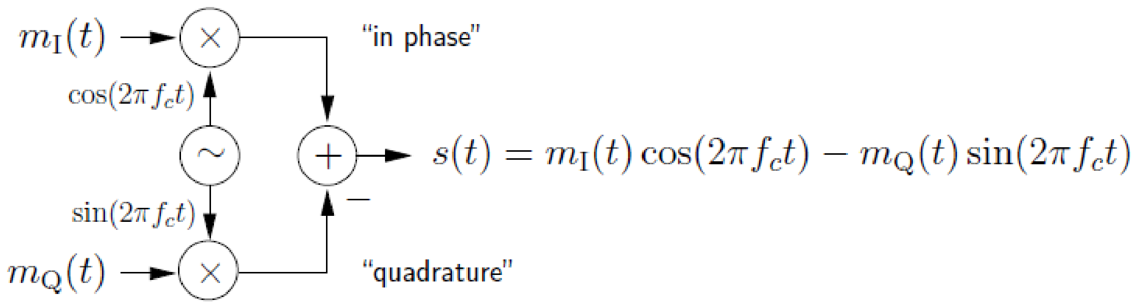
≪ MATLAB code here ≫



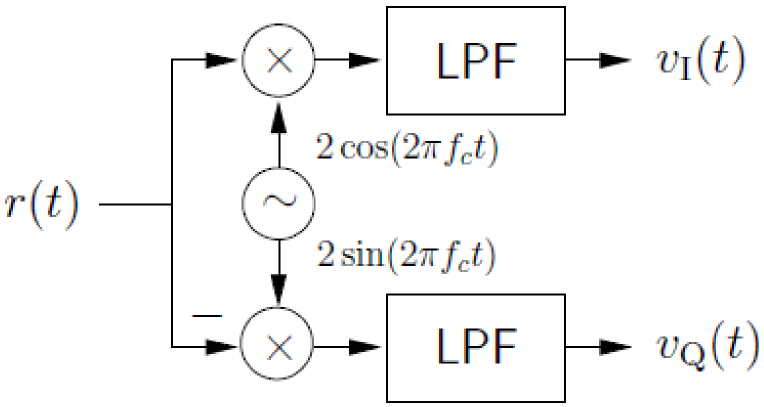
Quadrature Amplitude Modulation (QAM)

QAM is motivated by unwanted redundancy in the AM spectrum, which was symmetric around f_c.

QAM sends two real-valued signals $\{m_I(t), m_Q(t)\}$ simultaneously, resulting in a non-symmetric spectrum.



QAM demodulation is accomplished by:



where the LPF specs are the same as in AM, i.e., passband edge $B_p \geq W$ Hz and stopband edge $B_s \leq 2f_c - W$ Hz.

For a trivial channel, we have $r(t) = s(t)$, so that

Equation:

$$\begin{aligned}
v_I(t) &= \mathrm{LPF}\{r(t)\cdot 2\cos(2\pi f_c t)\} \\
&= \mathrm{LPF}\quad m_I(t)\underbrace{2\cos^2(2\pi f_c t)}_{1+\cos(4\pi f_c t)} \\
&\quad -m_{\mathrm{Q}}(t)\underbrace{2\sin(2\pi f_c t)\cos(2\pi f_c t)}_{\sin(4\pi f_c t)} \\
&= m_I(t) \\
v_{\mathrm{Q}}(t) &= \mathrm{LPF}\{-r(t)\cdot 2\sin(2\pi f_c t)\} \\
&= \mathrm{LPF}\quad -m_I(t)\underbrace{2\cos(2\pi f_c t)\sin(2\pi f_c t)}_{\sin(4\pi f_c t)} \\
&\quad +m_{\mathrm{Q}}(t)\underbrace{2\sin^2(2\pi f_c t)}_{1-\cos(4\pi f_c t)} \\
&= m_Q(t),
\end{aligned}$$

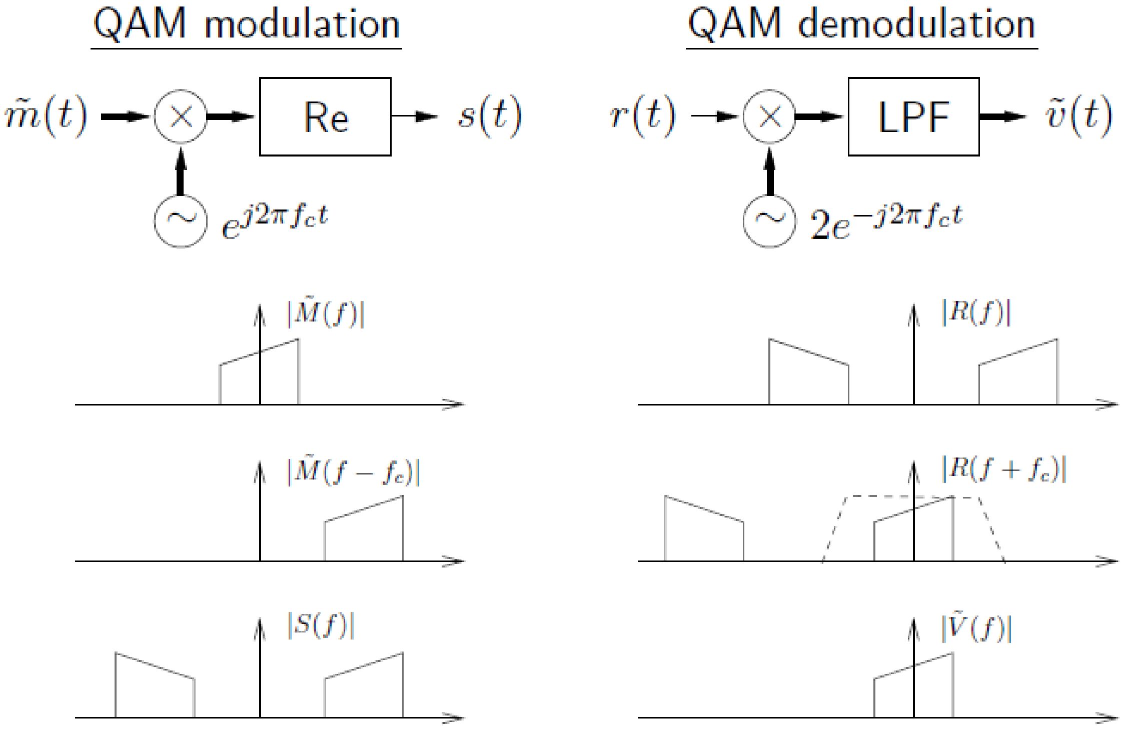
assuming synchronized oscillators.

When the oscillators are not synchronized, one gets coupling between the I&Q components as well as attenuation of each. Writing the I&Q signals in the “complex-baseband” form

Equation:

$$\begin{aligned}
\widetilde{m}(t) &= m_I(t) + j m_{\mathrm{Q}}(t) \\
\tilde{v}(t) &= v_I(t) + j v_{\mathrm{Q}}(t)
\end{aligned}$$

yields a much simpler description of QAM:



Note: Re $\{u(t)\} = \frac{1}{2}\left[u(t) + u^*(t)\right] \overset{\mathscr{F}}{\leftrightarrow} \frac{1}{2}\left[U(f) + U^*(-f)\right].$

We now verify the complex-baseband model for modulation:

Equation:

$$\begin{aligned}
&\text{Re}\left\{\tilde{m}(t)e^{j2\pi f_c t}\right\} \\
&= \text{Re}\left\{(m_I(t) + jm_Q(t))(\cos(2\pi f_c t) + j\sin(2\pi f_c t))\right\} \\
&= m_I(t)\cos(2\pi f_c t) - m_Q(t)\sin(2\pi f_c t) = s(t),
\end{aligned}$$

as well as for demodulation (assuming $r(t) = s(t)$):

Equation:

$$
\begin{aligned}
\tilde{v}(t) &= \mathrm{LPF}\left\{s(t)\cdot 2e^{-j2\pi f_c t}\right\} \\
&= \mathrm{LPF}\left\{\left(m_I(t)\cos(2\pi f_c t) - m_{\mathrm{Q}}(t)\sin(2\pi f_c t)\right)\cdot 2e^{-j2\pi f_c t}\right\} \\
&= \mathrm{LPF}\Big\{m_I(t)\left(e^{j2\pi f_c t} + e^{-j2\pi f_c t}\right)e^{-j2\pi f_c t} \\
&\quad -m_{\mathrm{Q}}(t)\left(je^{-j2\pi f_c t} - je^{j2\pi f_c t}\right)e^{-j2\pi f_c t}\Big\} \\
&= \mathrm{LPF}\left\{m_I(t)\left(1 + e^{-j4\pi f_c t}\right) - m_{\mathrm{Q}}(t)\left(je^{-j4\pi f_c t} - j\right)\right\} \\
&= m_I(t) + jm_{\mathrm{Q}}(t).
\end{aligned}
$$

The convenience of complex-baseband results in widespread use of complex-valued signals for comm systems!

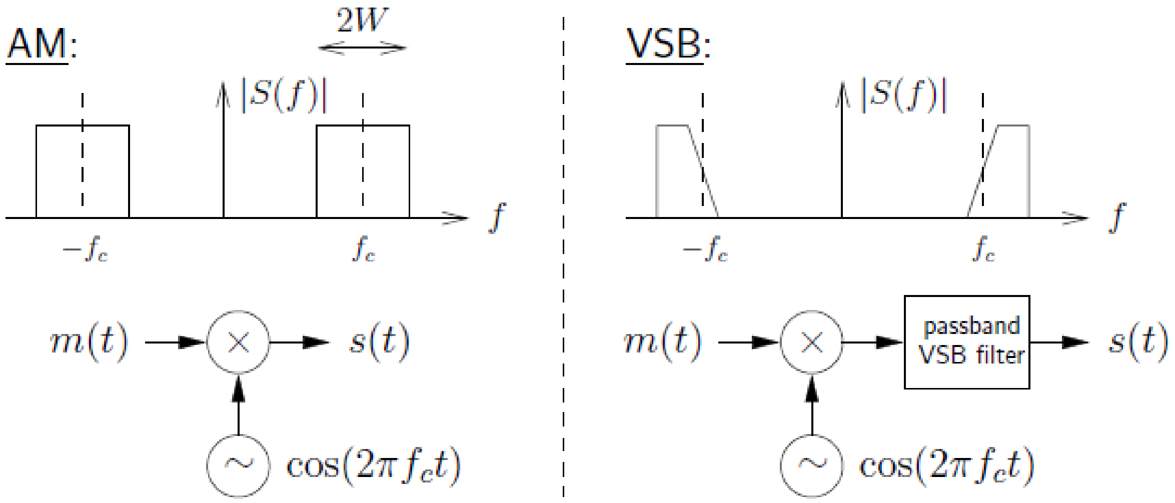
Note: To get the complex baseband formulation for AM, we simply set $m_{\mathrm{Q}}(t) = 0$ and $m_I(t) = m(t)$.

Vestigial Sideband Modulation (VSB)

VSB is another way to restore regain the spectral efficiency lost in AM. It's used to transmit North American terrestrial TV, both analog (NTSC) and digital (ATSC) formats.

Like AM, it can operate with or without a carrier tone.

Basically, VSB suppresses most of the redundant AM spectrum by filtering it:

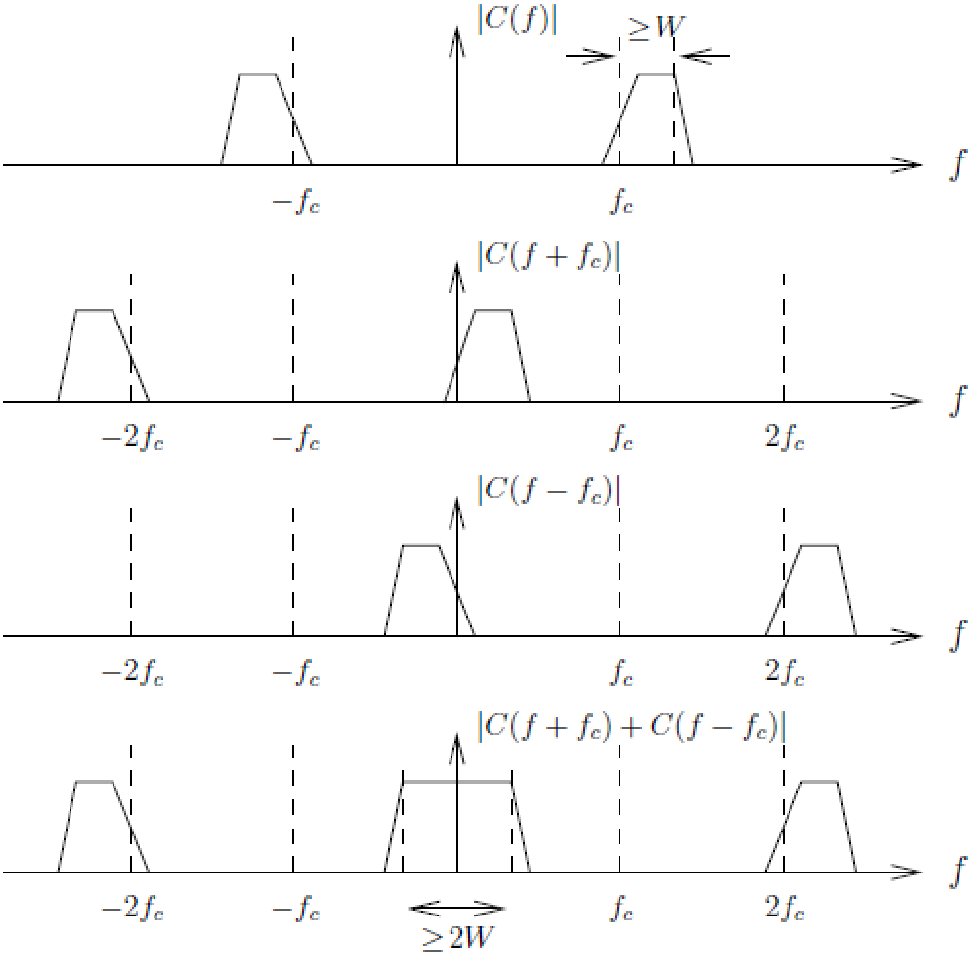


The passband VSB filter is a BPF $C(f)$ where

Equation:

$$C\left(f-f_c\right)+C\left(f+f_c\right)=2 \quad \text{for} \;\; |f| \leq W,$$

which implies its inside rolloff is symmetric around $f=f_c$:



For VSB modulation, we have

Equation:

$$\begin{aligned} s(t) &= m(t)\cos(2\pi f_c t) * c(t) \\ S(f) &= \frac{1}{2}[M(f+f_c)+M(f-f_c)]C(f). \end{aligned}$$

It turns out that VSB demod is *identical* to AM demod:

Equation:

$$\begin{aligned}
v(t) &= \text{LPF}\{r\,(t)\cdot 2\cos\,(2\pi f_c t)\} \\
&= \text{LPF}\{s\,(t)\cdot 2\cos\,(2\pi f_c t)\} \quad (\text{trivial channel}) \\
V(f) &= \text{LPF}\{S\,(f-f_c)+S\,(f+f_c)\} \\
&= \frac{1}{2}\text{LPF}\Big\{[M\,(f)+M\,(f-2f_c)]C\,(f-f_c) \\
&\quad +[M\,(f+2f_c)+M\,(f)]C\,(f+f_c)\} \\
&= M\,(f)\underset{=1 \text{ for } f\in[-W,W]}{\frac{1}{2}[C\,(f-f_c)+C\,(f+f_c)]} \\
&= M(f).
\end{aligned}$$

We note that the property

Equation:

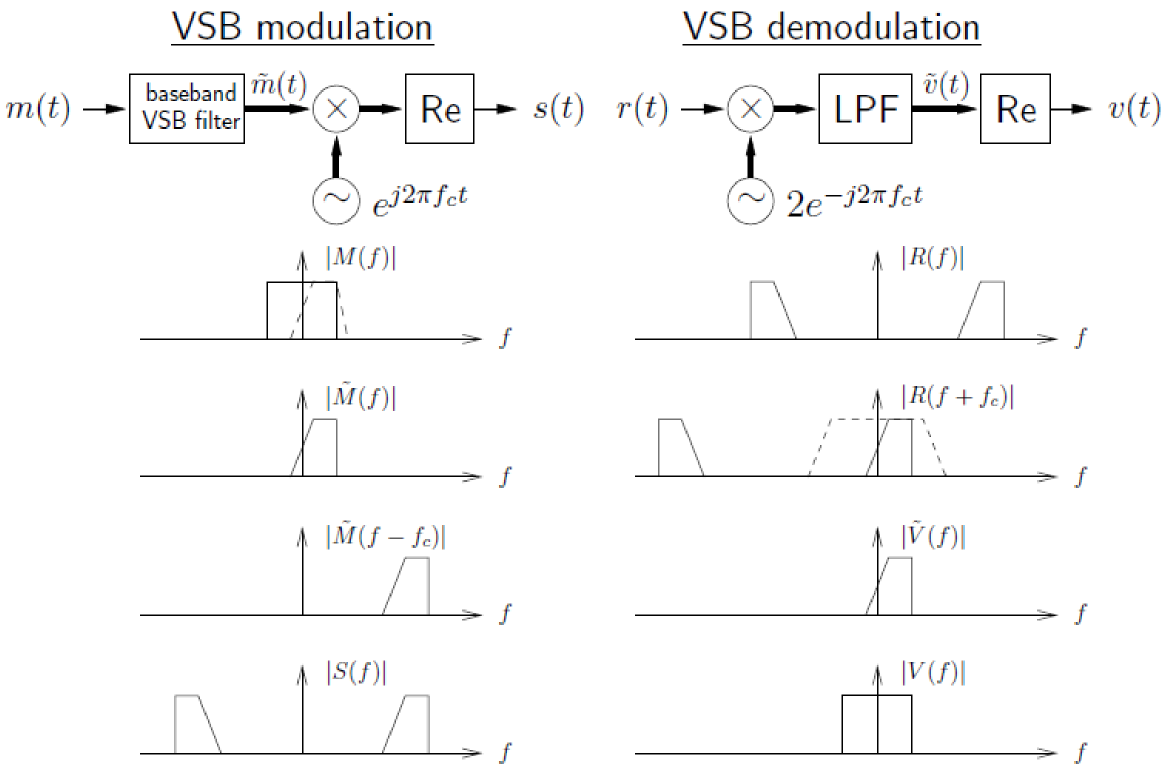
$$\mathscr{F}\{\cos\,(2\pi f_c t)c\,(t)\} \;=\; \frac{1}{2}[C\,(f-f_c)+C\,(f+f_c)]$$

may be convenient, e.g., for testing whether a given filter $c(t)$ satisfies the passband VSB criterion. VSB filtering can also be implemented at baseband using a complex-valued filter response $\tilde{c}(t)$ which satisfies

Equation:

$$\widetilde{C}\,(f)+\widetilde{C}^*\,(-f)=2 \quad \text{for} \;\; |f|\le W,$$

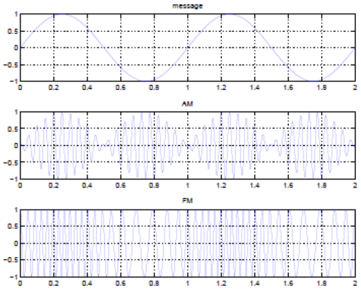
generating the complex-baseband message signal $\widetilde{m}\,(t)$. The message can be recovered by simplying ignoring the imaginary part of the complex-baseband output $\tilde{v}\,(t)$.



Motivation: filtering at baseband is usually much cheaper than filtering at passband.

Frequency Modulation (FM)

While AM modulated the carrier amplitude, FM modulates the carrier frequency.



```
t_max = 2.0; W = 1;          % message
params
Ts = 1/1000; t = 0:Ts:t_max;
m = sin(2*pi*W*t);           % message
signal
fc = 20;                     % carrier
freq
D = 15;                      % FM mod
```

```
index
kf = D*W/max(abs(m));      % freq
sensitivity
s_am = m.*cos(2*pi*fc*t);
s_fm =
cos(2*pi*fc*t+2*pi*kf*cumsum(m)*Ts);
subplot(3,1,1)
plot(t,m);
grid on; title('message');
subplot(3,1,2)
plot(t,s_am);
grid on; title('AM');
subplot(3,1,3)
plot(t,s_fm);
grid on; title('FM');
```

In particular, FM modulates the real-valued message $m(t)$ via

Equation:

$$s\left(t\right) = \cos \quad 2\pi f_c t + \qquad 2\pi k_{\mathrm{f}} \int_0^t m\left(\tau\right) d\tau \qquad .$$

$$\phi(t)\text{"instantaneous modulation phase"}$$

where k_{f} is called the “frequency-sensitivity factor.” Since the instantaneous modulation frequency

Equation:

$$\frac{d\phi(t)}{dt} = 2\pi k_{\mathrm{f}}\ m\left(t\right)$$

is a scaled version of the message $m(t)$, it is fitting to call this scheme “frequency modulation.”

Using the peak frequency deviation $\Delta_{\mathrm{f}} = k_{\mathrm{f}} \max |m(t)|$, the “modulation index” D is defined as

Equation:

$$D \;=\; \frac{\Delta_{\mathrm{f}}}{W} \leftarrow \text{denominator is the single-sided BW of } m(t).$$

Increasing D decreases spectral efficiency but increases robustness to noise/interference.

Equation:

$$\begin{aligned} D \ll 1 &\;:\; \text{``narrowband FM''}, \\ D \gg 1 &\;:\; \text{``wideband FM''}. \end{aligned}$$

Carson's Rule approximates the FM passband signal-BW as

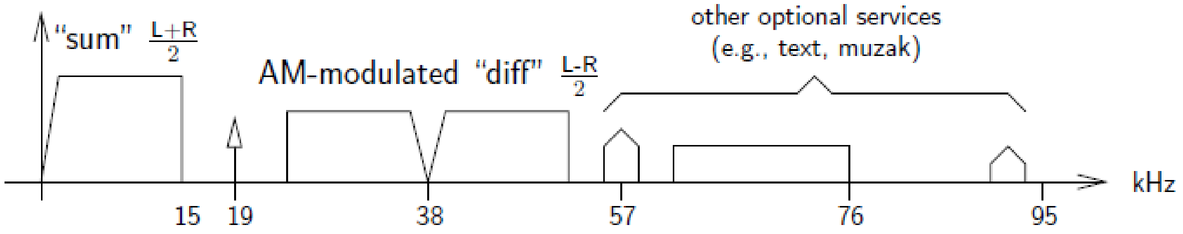
Equation:

$$\mathrm{BW}_{99} \;\approx\; 2\,(\Delta_{\mathrm{f}} + W) = 2\,(D+1)W.$$

Example: Mono FM radio:

- Message signal filtered to freq interval [30,15k] Hz.
- FCC limits $\Delta_{\mathrm{f}} \leq 75$ kHz (channels 200 kHz apart). $\Rightarrow D = \frac{75}{15} = 5$

FM stereo uses smaller D due to message spectrum:



There are various FM demodulators, but the “discriminator” is one of the best known. Recalling that

Equation:

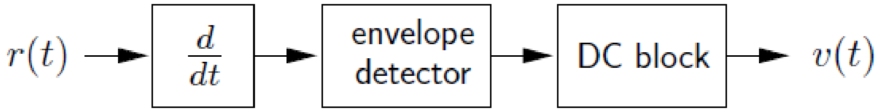
$$\frac{d}{dt}\cos\left(\phi(t)\right) = -\frac{d\phi(t)}{dt}\sin\left(\phi(t)\right),$$

we see that

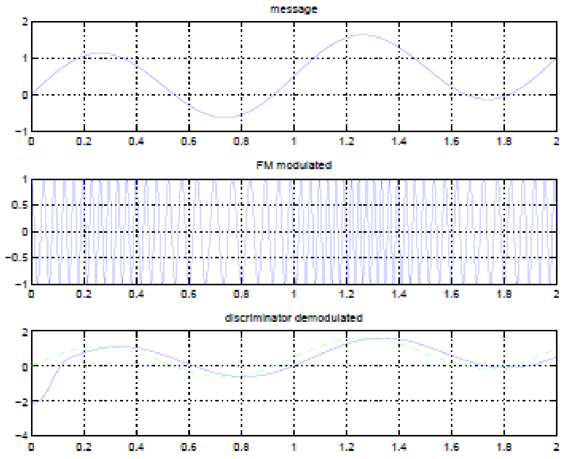
Equation:

$$\begin{aligned}\frac{d}{dt}s\left(t\right) &= \tfrac{d}{dt}\cos\left(2\pi f_c t + 2\pi k_{\mathrm{f}}\int_0^t m\left(\tau\right)d\tau\right)\\ &= -\left[2\pi f_c + 2\pi k_{\mathrm{f}}\, m\left(t\right)\right]\sin\left(2\pi f_c t + 2\pi\int_0^t m\left(\tau\right)d\tau\right)\end{aligned}$$

is a form of large-carrier AM (assuming $f_c > k_{\mathrm{f}}\, m\left(t\right)$), which can be demodulated using an envelope detector as follows:



≪ MATLAB code here ≫



Random Signals and Noise

This module describes the basics of random signals and noise as needed for an introductory course on data communications. First it introduces the notion of power spectral density (PSD) and defines "white" noise as a noise with a flat PSD. Next it describes the effect that linear filtering has on a random signal, in both time and frequency domains. For this, the autocorrelation function is introduced, and expectations are used to derive the fundamental properties.

When designing a comm system, it is impossible to know exactly what the signal and noise waveforms will be. But usually we know average characteristics such as energy distribution across frequency. It is exactly these statistics that are most often used for comm system design.

We will consider random waveforms known as “zero-mean wide-sense stationary random processes.” Such an $x(t)$ is completely completely characterized by its *power spectral density* $S_x(f)$, or average signal power versus frequency *f*. The technical definition of PSD accounts for the fact that power is defined as energy per unit time:

Equation:

$$S_x(f) := \lim_{T\to\infty} \frac{1}{2T} \operatorname{E}\left\{\left|\int_{-T}^{T} x(t)e^{j2\pi ft}dt\right|^2\right\},$$

Above, $\operatorname{E}\{\cdot\}$ denotes *expectation*, i.e., statistical average.

A common random waveform is “white noise.” Saying that $w(t)$ is *white* is equivalent to saying that its PSD is flat:

Equation:

$$S_w(f) = N_o \text{ for all } f.$$

Broadband noise is often modelled as white random noise. A fundamentally important question is: **What does filtering do to the power spectrum of a signal?** The answer comes with the aid of the *autocorrelation function* $R_x(\tau)$, defined as

Equation:

$$R_x(\tau) := \mathrm{E}\left\{x(t)x^*(t-\tau)\right\}$$

and having the important property

Equation:

$$R_x(\tau) \overset{\mathscr{F}}{\leftrightarrow} S_x(f).$$

Note that, for white noise $w(t)$ with PSD N_o, we have

Equation:

$$\begin{aligned} S_w(f) = N_o \quad &\Longrightarrow \quad R_w(\tau) = N_o\delta(\tau) \\ &\Longrightarrow \quad \mathrm{E}\left\{w(t)w^*(t-\tau)\right\} = N_o\delta(\tau). \end{aligned}$$

Here we see that $w(t_1)$ and $w(t_2)|t_2 \neq t_1$ are "uncorrelated" because $\mathrm{E}\left\{w(t_1)w^*(t_2)\right\} = 0$, and that $w(t)$ has average energy $\mathscr{E}_w = \mathrm{E}\{|w(t)|^2\} = \infty$, which may be surprising. We could have found the same via $\mathscr{E}_w = \int_{-\infty}^{\infty} S_w(f)df$.

Now say that white noise $w(t)$ is filtered with non-random $h(t)$, yielding output $y(t) = \int_{-\infty}^{\infty} w(q)h(t-q)dq$. We can find $S_y(f)$ by first finding $R_y(\tau)$ and then taking the FT. Recall:

Equation:

$$R_y(\tau) = \mathrm{E}\left\{y(t)y^*(t-\tau)\right\}.$$

Plugging in the expression for $y(t)$, we find

Equation:

$$\begin{aligned} R_y(\tau) &= \mathrm{E}\left\{\int_{-\infty}^{\infty} w(q_1)h(t-q_1)dq_1 \int_{-\infty}^{\infty} w^*(q_2)h^*(t-\tau-q_2)dq_2\right\} \\ &= \mathrm{E}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(q_1)w^*(q_2)\,h(t-q_1)h^*(t-\tau-q_2)dq_1dq_2\right\}. \end{aligned}$$

Because averaging is a linear operation, and because the $h(\cdot)$ terms are fixed (non-random) quantities, we can write

Equation:

$$R_y(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{E}\left\{w(q_1)w^*(q_2)\right\} h(t-q_1)h^*(t-\tau-q_2)dq_1dq_2.$$

Since $w(t)$ is white,

Equation:

$$\mathrm{E}\left\{w(q_1)w^*(q_2)\right\} = \mathrm{E}\left\{w(q_1)w^*(q_1-(q_1-q_2))\right\} = N_o\delta(q_1-q_2),$$

which allows use of the sifting property on the inner integral:

Equation:

$$\begin{aligned} R_y(\tau) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N_o\delta(q_1-q_2)h(t-q_1)h^*(t-\tau-q_2)dq_1dq_2 \\ &= N_o\int_{-\infty}^{\infty} h(t-q_2)h^*(t-\tau-q_2)dq_2 \\ &= N_o\int_{\infty}^{-\infty} h(q)h^*(q-\tau)(-dq) \quad \text{using } q := t-q_2 \\ &= N_o\int_{-\infty}^{\infty} h(q)h^*(q-\tau)dq. \end{aligned}$$

To summarize, the autocorrelation of filtered white noise is

Equation:

$$R_y(\tau) = N_o \int_{-\infty}^{\infty} h(q) h^*(q-\tau) dq.$$

Having $R_y(\tau)$, the final step is finding $S_y(f) = \mathscr{F}\{R_y(\tau)\}$:

Equation:

$$\begin{aligned}
S_y(f) &= \int_{-\infty}^{\infty} R_y(\tau) e^{-j2\pi f\tau} d\tau \\
&= \int_{-\infty}^{\infty} \left[N_o \int_{-\infty}^{\infty} h(q) h^*(q-\tau) dq \right] e^{-j2\pi f\tau} d\tau \\
&= N_o \int_{-\infty}^{\infty} h(q) e^{-j2\pi fq} \left[\int_{-\infty}^{\infty} h^*(q-\tau) e^{j2\pi f(q-\tau)} d\tau \right] dq \\
&= N_o \int_{-\infty}^{\infty} h(q) e^{-j2\pi fq} \left[\int_{-\infty}^{\infty} h^*(t) e^{j2\pi ft} dt \right] dq \\
&= N_o \int_{-\infty}^{\infty} h(q) e^{-j2\pi fq} dq \left[\int_{-\infty}^{\infty} h(t) e^{-j2\pi ft} dt \right]^* \\
&= N_o H(f) H^*(f) \ = \ N_o |H(f)|^2
\end{aligned}$$

To summarize, the power spectrum of filtered white noise is

Equation:

$$S_y(f) = N_o |H(f)|^2.$$

More generally, the power spectrum of a filtered random process $x(t)$ can be shown to be

Equation:

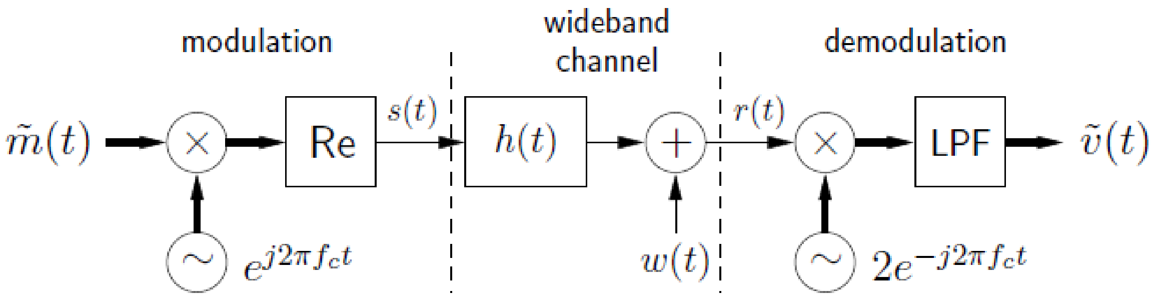
$$S_y(f) = S_x(f) \, |H(f)|^2,$$

which is quite intuitive.

Complex-Baseband Equivalent Channel

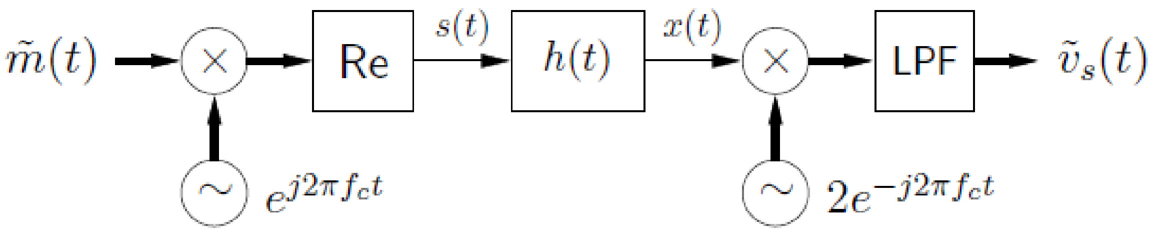
In this module, the "complex-baseband equivalent channel" is derived from the three-part cascade of modulation, wideband channel, and demodulation.

Linear communication schemes (e.g., AM, QAM, VSB) can all be represented (using complex-baseband mod/demod) as:



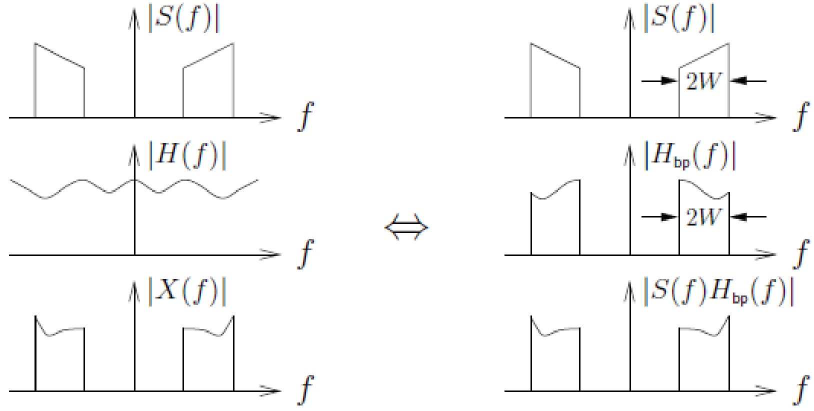
It turns out that this diagram can be greatly simplified...

First, consider the signal path on its own:



Since $s(t)$ is a bandpass signal, we can replace the *wideband* channel response $h(t)$ with its *bandpass equivalent* $h_{\text{bp}}(t)$:

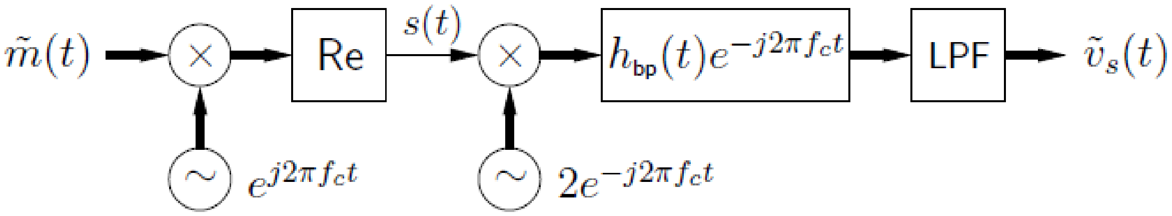
...because filtering $s(t)$ with $h(t)$ is equivalent to filtering $s(t)$ with $h_{\text{bp}}(t)$:



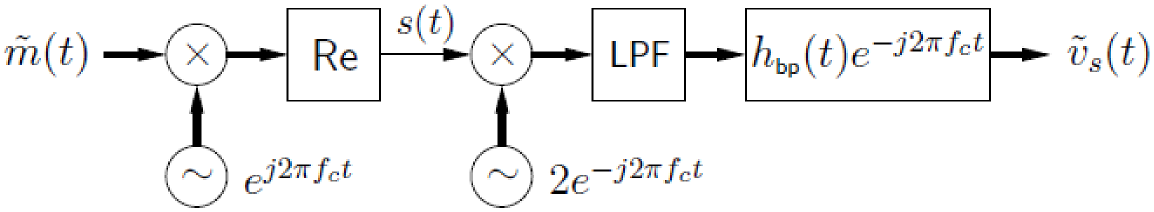
Then, notice that
Equation:

$$\begin{aligned}[s(t)*h_{\text{bp}}(t)]2e^{-j2\pi f_c t} &= \int s(\tau)h_{\text{bp}}(t-\tau)d\tau \cdot 2e^{-j2\pi f_c t} \\ &= \int s(\tau)2e^{-j2\pi f_c \tau}h_{\text{bp}}(t-\tau)e^{-j2\pi f_c(t-\tau)}d\tau \\ &= \left[s(t)2e^{-j2\pi f_c t}\right]*\left[h_{\text{bp}}(t)e^{-j2\pi f_c t}\right],\end{aligned}$$

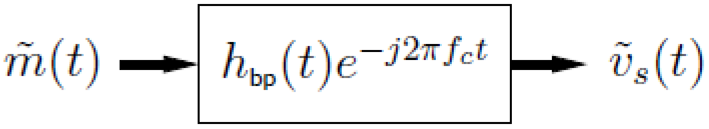
which means we can rewrite the block diagram as



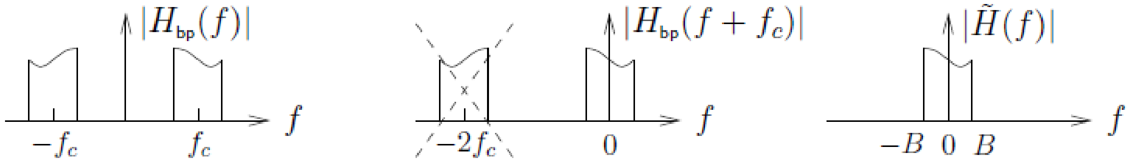
We can now reverse the order of the LPF and $h_{\text{bp}}(t)e^{-j2\pi f_c t}$ (since both are LTI systems), giving



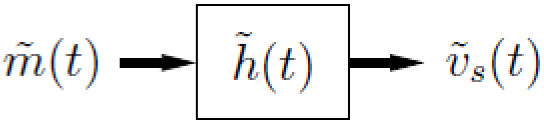
Since mod/demod is transparent (with synched oscillators), it can be removed, simplifying the block diagram to



Now, since $\tilde{m}(t)$ is bandlimited to W Hz, there is no need to model the left component of $H_{\text{bp}}(f+f_c) = \mathscr{F}\left\{h_{\text{bp}}(t)e^{-j2\pi f_c t}\right\}$:

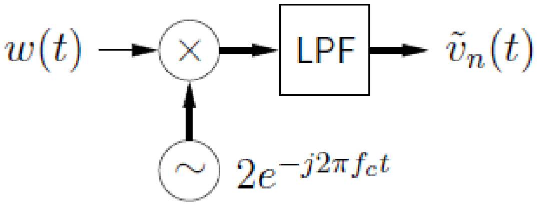


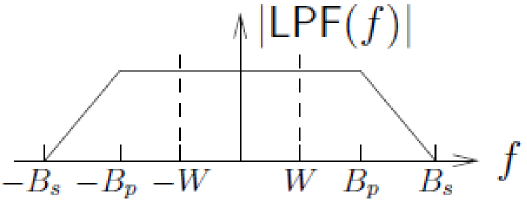
Replacing $h_{\text{bp}}(t)e^{-j2\pi f_c t}$ with the *complex-baseband response* $\tilde{h}(t)$ gives the “complex-baseband equivalent” signal path:

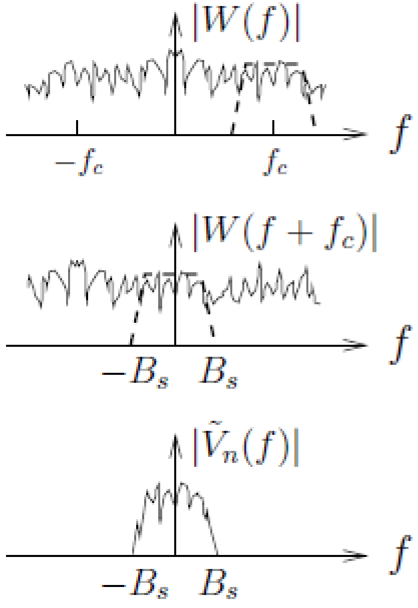


The spectrums above show that $h_{\text{bp}}(t) = \text{Re}\left\{\tilde{h}(t) \cdot 2e^{j2\pi f_c t}\right\}$.

Next consider the noise path on it's own:

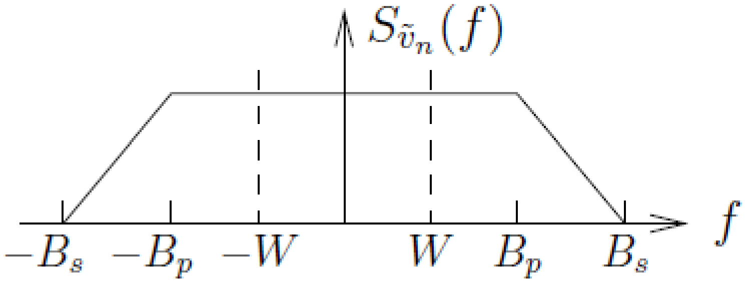




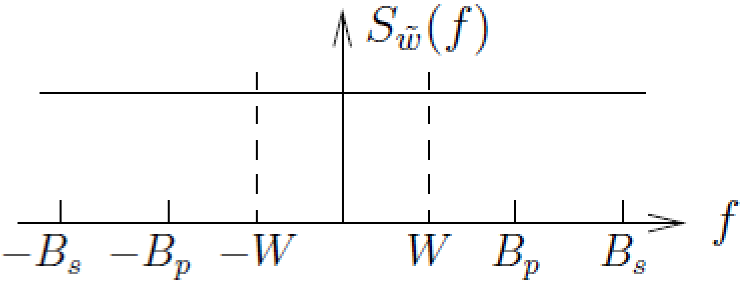


From the diagram, $\tilde{v}_n(t)$ is a baseband version of the bandpass noise spectrum that occupies the frequency range $f \in [f_c - B_s, f_c + B_s]$. Since $\tilde{v}_n(t)$ is complex-valued, $\widetilde{V}_n(f)$ is non-symmetric.

Say that $w(t)$ is real-valued white noise with power spectral density (PSD) $S_w(f) = N_0$. Since $S_w(f)$ is constant over all *f*, the PSD of the complex noise $\tilde{v}_n(t)$ will be constant over the LPF passband, i.e., $f \in [-B_p, B_p]$:

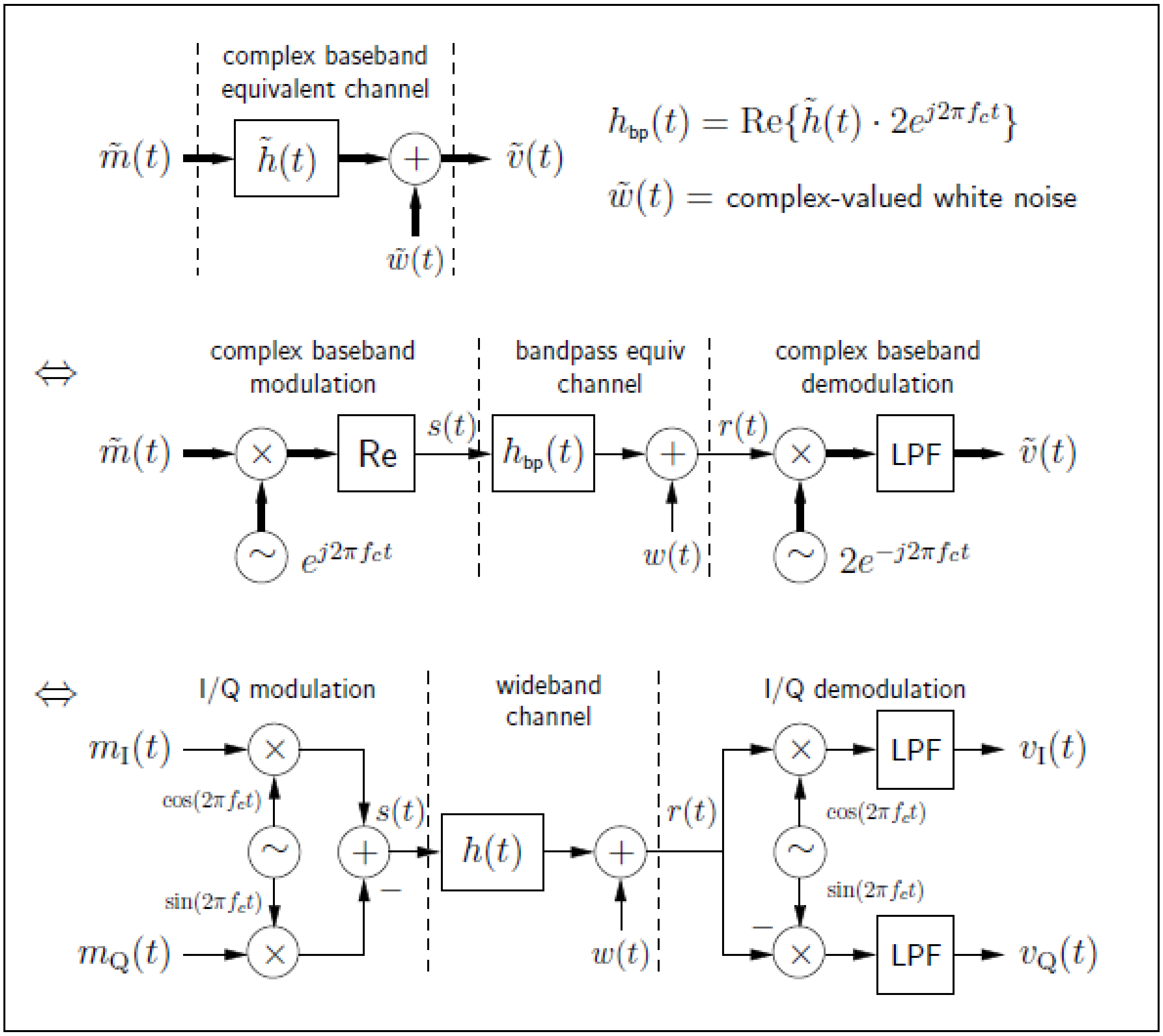


A well-designed communications receiver will suppress all energy outside the signal bandwidth *W*, since it is purely noise. Given that the noise spectrum outside $f \in [-W, W]$ will get totally suppressed, *it doesn't matter how we model it!* Thus, we choose to replace the lowpass complex noise $\tilde{v}_n(t)$ with something simpler to describe: *white* complex noise $\widetilde{w}(t)$ with PSD $S_{\widetilde{w}}(f) = N_0$:



We'll refer to $\widetilde{w}(t)$ as "*complex baseband equivalent*" noise.

Putting the signal and noise paths together, we arrive at the *complex baseband equivalent channel model*:



The diagrams above should convince you of the utility of the complex-baseband representation in simplifying the system model!

Digital Communication

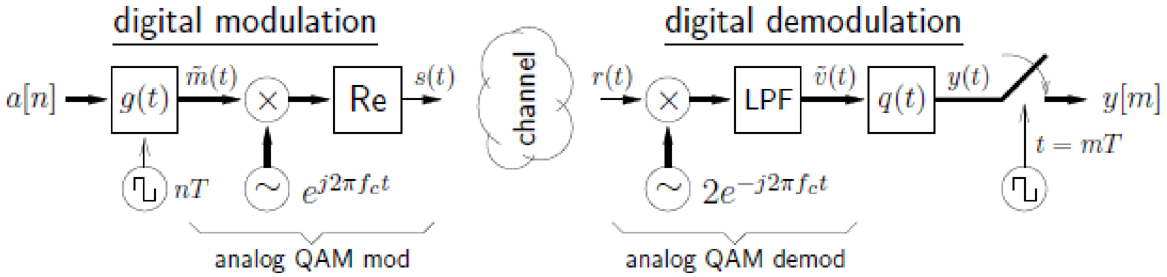
This module outlines the basics of digital communication through a white-noise channel. Transmission consists of pulse shaping and modulation, while reception consists of demodulation, filtering, and sampling. The combined pulse-shaping and filtering designs which prevent inter-symbol interference (ISI) are identified as those satisfying the Nyquist criterion, which is examined in both the time and frequency domains, and the raised-cosine (combined) pulse is given as a common example. The Cauchy-Schwarz inequality is then used to show the matched filtering maximizes the signal-to-noise ratio. Finally, the square-root raised cosine pulse/filter is given as an example of matched filtering that satisfies the Nyquist criterion.

Transmission consists of

1. pulse shaping: $\widetilde{m}(t) = \sum_n a[n] g(t - nT)$,
2. modulation: $s(t) = \mathrm{Re}\left\{\widetilde{m}(t) e^{j2\pi f_c t}\right\}$.

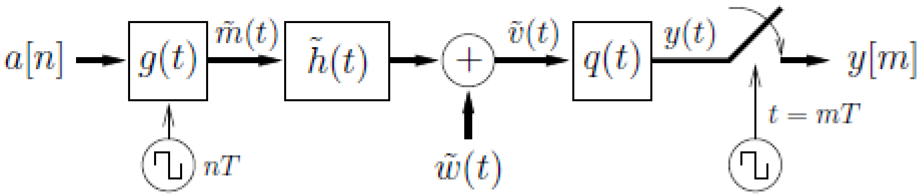
Reception consists of

1. demodulation: $\tilde{v}(t) = \mathrm{LPF}\left\{2r(t) e^{-j2\pi f_c t}\right\}$,
2. filtering: $y(t) = \tilde{v}(t) * q(t)$,
3. sampling: $y[m] = y(mT)$.



Building on analog QAM mod/demod components, digital mod adds pulse shaping & demod adds filtering/sampling.

Simplifying via the complex-baseband equivalent channel:



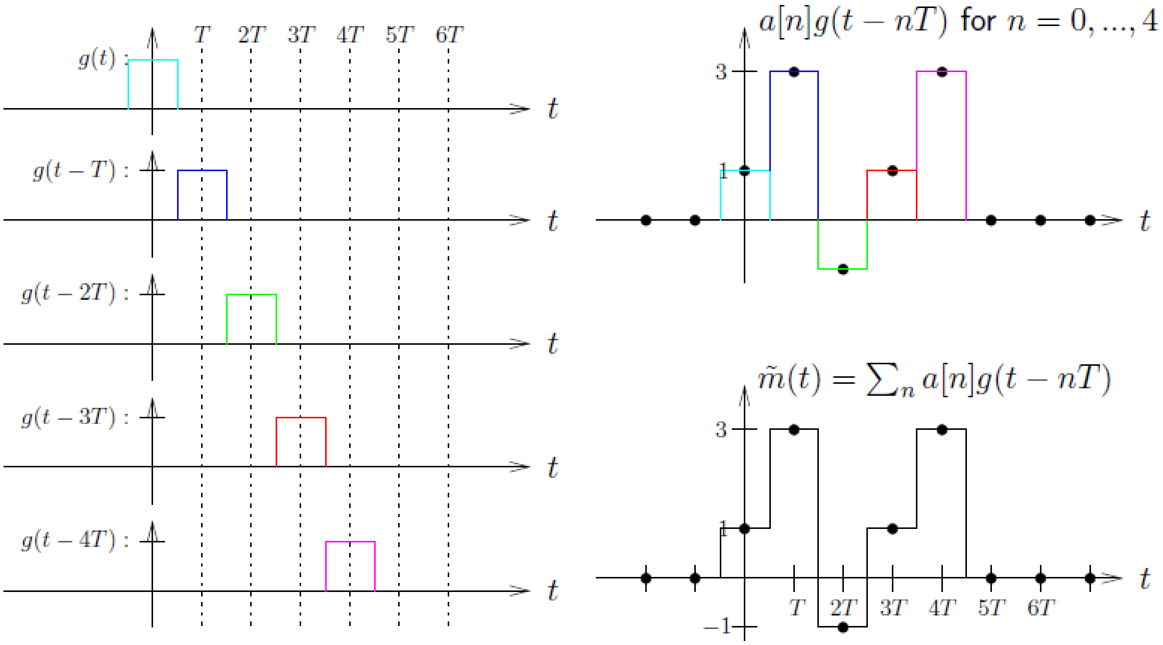
Transmitter pulse shaping is used to convert the symbol sequence $\{a[n]\}$ into the continuous message $\tilde{m}(t)$:

Equation:

$$\begin{aligned} \tilde{m}(t) &= \sum_n a[n] g(t - nT) \quad \text{"baseband message"} \\ T &= \text{"symbol period"} \end{aligned}$$

Thus, $\tilde{m}(t)$ can be seen to be a superposition of *scaled* and *time-shifted* copies of the pulse waveform $g(t)$.

Example, if the symbol sequence $[a[0], a[1], a[2], a[3], a[4]]$ equals $[1, 3, -1, 1, 3]$, then the square pulse $g(t)$ shown below left yields the message $\tilde{m}(t)$ shown below right.



Receiver filtering (via $q(t)$) has two goals:

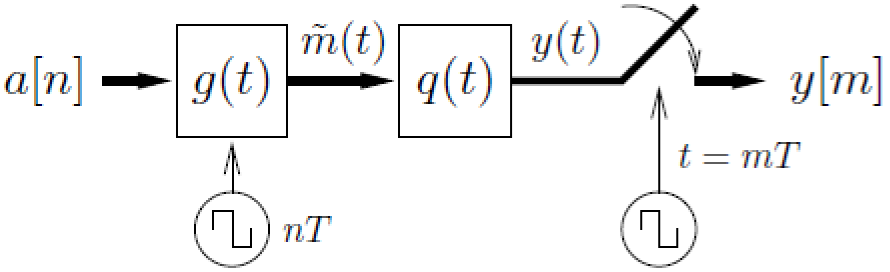
1. noise suppression (i.e., SNR improvement),
2. inter-symbol interference (ISI) prevention.

Noise suppression was briefly discussed in Preliminaries and will soon be revisited in more detail. Next we describe ISI.

Realize that, in the *ideal* digital comm system, the n^{th} output $y[n]$ would simply equal the n^{th} input $a[n]$. But in practice, $y[n]$ can be corrupted by interference from the other symbols $\{a\,[m]\}_{m\neq n}$, known as "inter-symbol interference," and noise.

ISI-prevention for the noiseless trivial channel

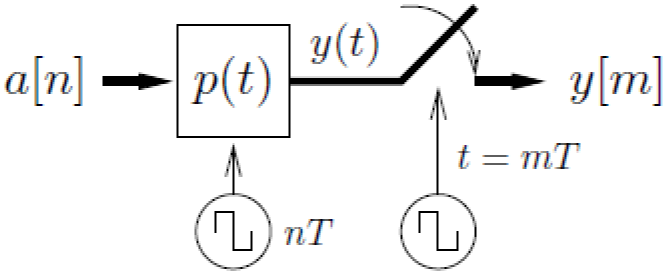
Consider the idealized system



Equation:

$$\begin{aligned} y(t) &= \int q\,(\tau)\tilde{m}\,(t-\tau)d\tau \quad \text{for} \quad \tilde{m}\,(t) = \textstyle\sum_n a\,[n]g\,(t-nT) \\ &= \sum_n a\,[n] \int q\,(\tau)g\,(t-nT-\tau)d\tau \\ &= \sum_n a\,[n]p\,(t-nT) \quad \text{for} \quad p\,(t) = g\,(t)*q\,(t). \end{aligned}$$

Thus, the idealized system can be re-drawn as

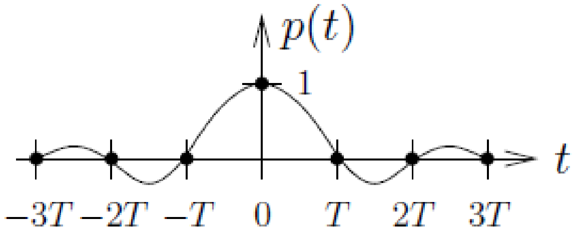


where

Equation:

$$y[m] = y(mT) = \sum_n a[n] p(mT - nT) = \sum_n a[n] p((m-n)T).$$

To make $y[m] = a[m]$ (i.e., prevent ISI), we need



$$p(mT) = \begin{cases} 1 & m = 0 \\ 0 & m \neq 0 \end{cases}$$

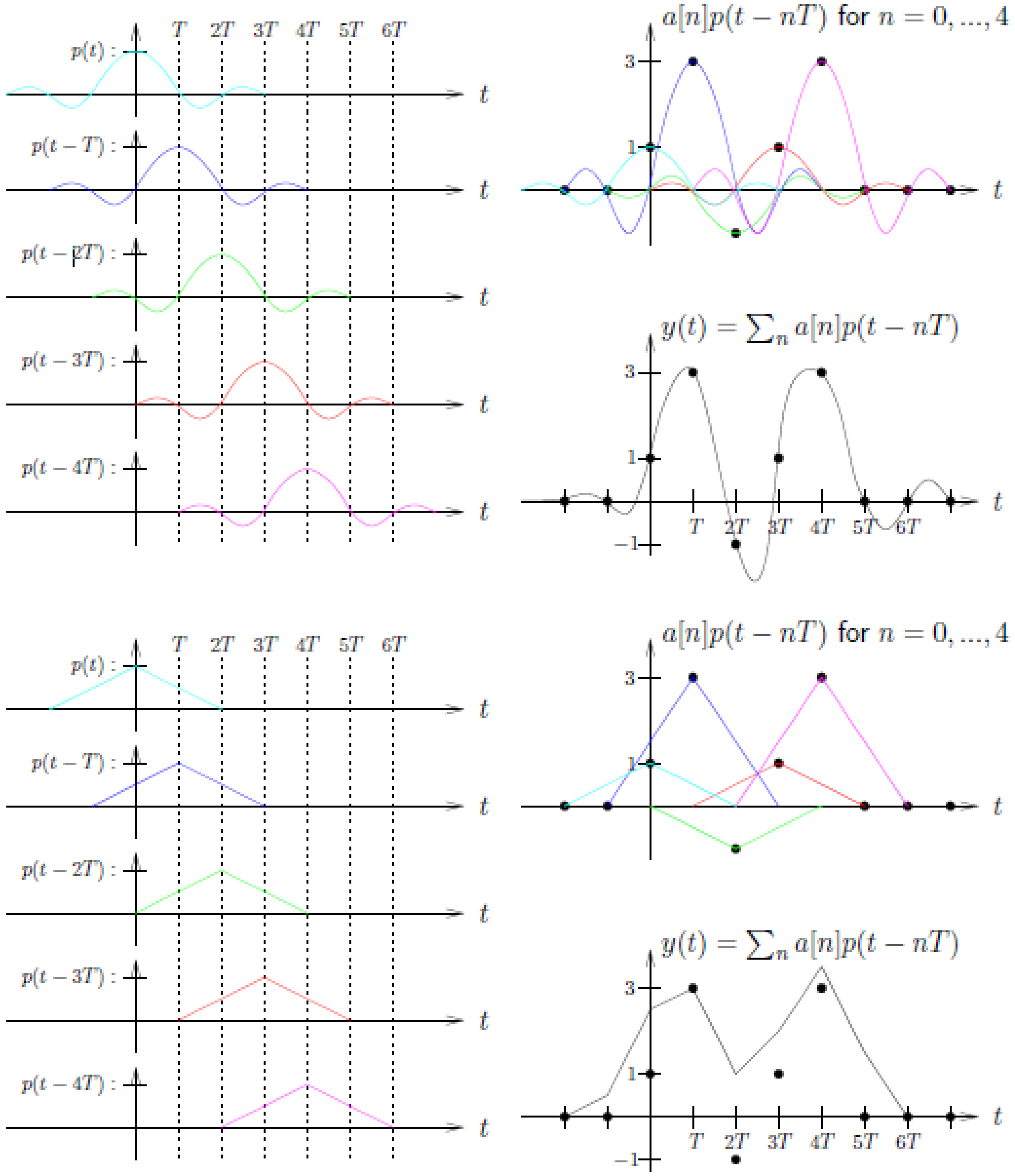
which is known as the “Nyquist Criterion.” This criterion can be simply stated as $p(mT) = \delta[m]$ using

Equation:

$$\delta\left[m\right] = \begin{cases} 1 & m = 0 \quad \text{"distcrete-time impulse,"} \\ 0 & m \neq 0 \quad \text{or "Kronecker delta."} \end{cases}$$

$$a[n] \;=\; \sum_{m=-\infty}^{\infty} a\,[m]\delta\,[n-m] \qquad \text{"sifting property."}$$

Examples of Nyquist, and non-Nyquist, combined-pulses $p(t)$ for $[a[0], a[1], a[2], a[3], a[4]] = [1, 3, -1, 1, 3]$:



There is an interesting frequency-domain interpretation. Since

Equation:

$$\frac{1}{T}\sum_{k=-\infty}^{\infty}\delta\left(f-\frac{k}{T}\right)\ \overset{\mathscr{F}}{\leftrightarrow}\ \sum_{m=-\infty}^{\infty}\delta\left(t-mT\right),$$

we can see that

Equation:

$$\underbrace{P(f)*\frac{1}{T}\sum_{k=-\infty}^{\infty}\delta\left(f-\frac{k}{T}\right)}_{\frac{1}{T}\sum_{k=-\infty}^{\infty}P\left(f-\frac{k}{T}\right)} \overset{\mathscr{F}}{\leftrightarrow} \underbrace{p(t)\cdot\sum_{m=-\infty}^{\infty}\delta(t-mT)}_{\sum_{m=-\infty}^{\infty}p(mT)\delta(t-mT)}.$$

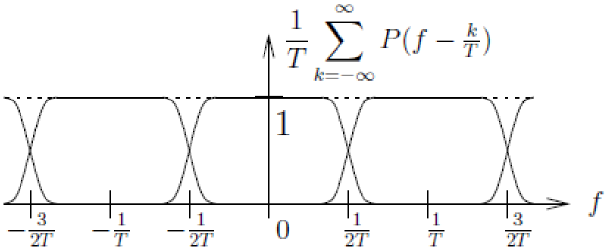
So, the time-domain Nyquist criterion $p(mT)=\delta[m]$ implies

Equation:

$$\frac{1}{T}\sum_{k=-\infty}^{\infty}P\left(f-\frac{k}{T}\right) \overset{\mathscr{F}}{\leftrightarrow} \delta(t),$$

which in turn implies

$$\frac{1}{T}\sum_{k=-\infty}^{\infty}P\left(f-\frac{k}{T}\right)=1.$$



In other words, the superposition of $\left\{\frac{1}{T}P\left(f-\frac{k}{T}\right)\right\}_{k\in\mathbb{Z}}$ must sum to one. This frequency-domain version of the Nyquist Criterion will soon come in handy...

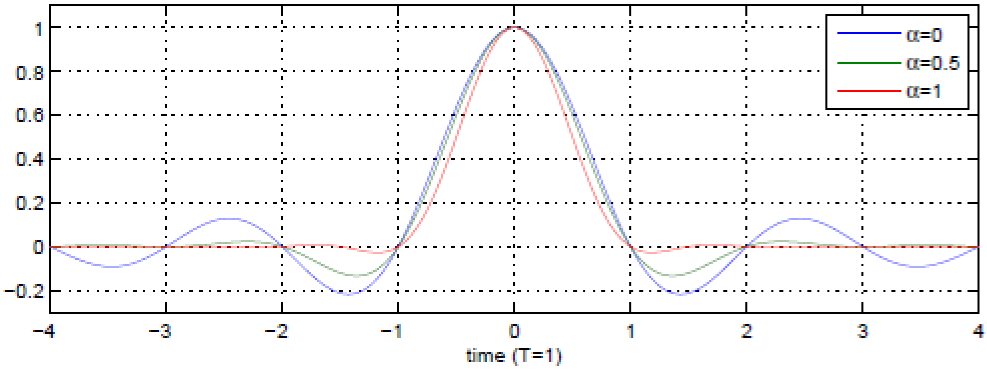
A popular choice of combined pulse $p(t)=g(t)*q(t)$ is the “raised-cosine pulse” with rolloff parameter $\alpha\in[0,1]$:

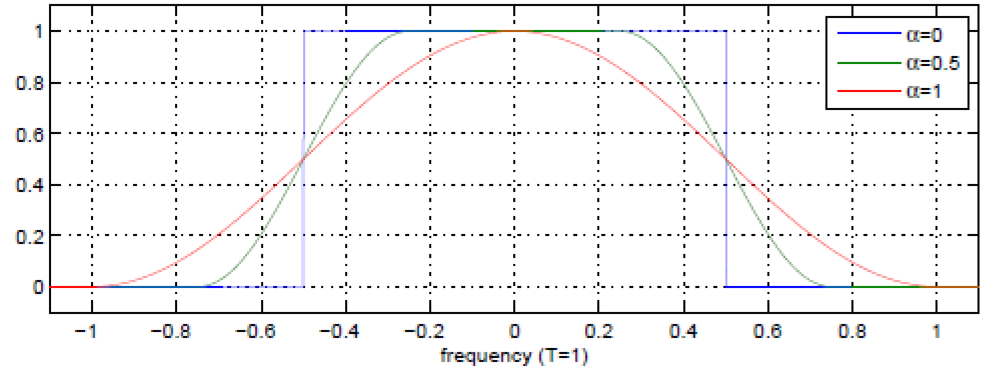
Equation:

$$p_{\mathrm{RC}}(t) \quad = \quad \frac{\cos(\pi\alpha t/T)}{1-(2\alpha t/T)^2}\ \mathrm{sinc}(t/T), \quad \mathrm{sinc}(x) := \frac{\sin(\pi x)}{\pi x}$$

$$P_{\mathrm{RC}}(f) \quad = \quad \begin{cases} T & |f| \le \frac{(1-\alpha)}{2T} \\ T\cos^2\left(\frac{\pi T}{2\alpha}\left(|f| - \frac{1-\alpha}{2T}\right)\right) & \frac{(1-\alpha)}{2T} \le |f| \le \frac{(1+\alpha)}{2T} \\ 0 & \frac{(1+\alpha)}{2T} \le |f| \end{cases}.$$

Tradeoff: larger $\alpha \;\Rightarrow$ less time-spread but more freq-spread:



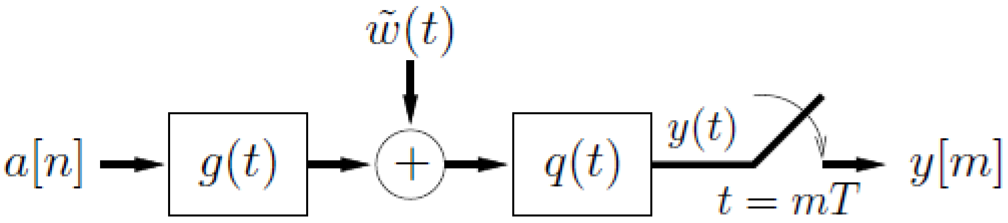


So, we now know how to design the combined pulse $p(t)$.

But what about the individual pulses $g(t)$ and $q(t)$?

Maximizing SNR for ISI-free Pulses in White Noise

Now let's bring the noise back into consideration. Given

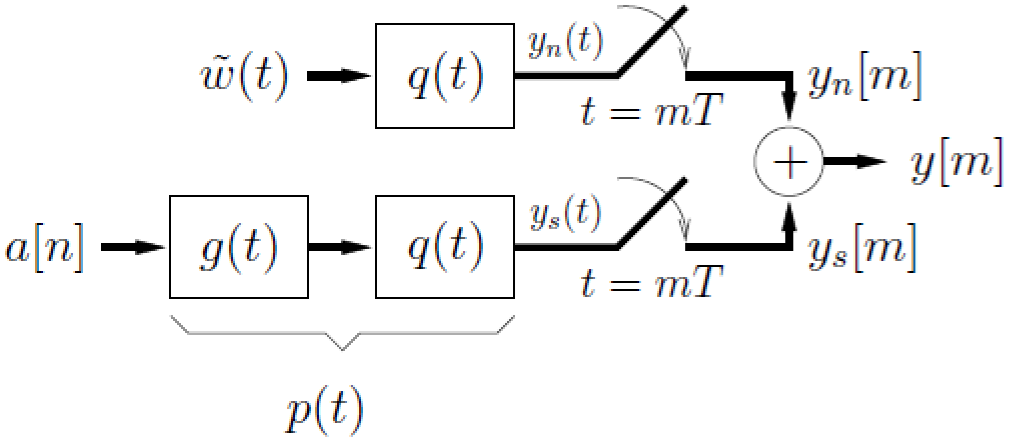


Equation:

$$\mathrm{E}\left\{\tilde{w}(t)\,\tilde{w}^*(t-\tau)\right\} = N_0\delta(\tau) \quad \text{(complex white noise)}$$

we want a $\{g(t), q(t)\}$ pair that maximizes the SNR of $y[m]$.

Separating the noise and signal contributions to $y[m]$ via



the SNR can be written

Equation:

$$\text{SNR} = \frac{\mathscr{E}_s}{\mathscr{E}_n} = \frac{\text{E}\left\{|y_s[m]|^2\right\}}{\text{E}\left\{|y_n[m]|^2\right\}},$$

where E_s and E_n are average signal and noise energies.

Here, we treat both $\widetilde{w}(t)$ and $a[n]$ as random, implying that $y_s[m]$ and $y_n[m]$ are both random.

Notice that, with an ISI-free combined pulse $p(t)$, we get
Equation:

$$\begin{aligned} y_s[m] &= \sum_n a[n]p((m-n)T) = a[m]p(0) \\ p(0) &= \int_{-\infty}^{\infty} q(\tau)g(0-\tau)d\tau, \end{aligned}$$

so that
Equation:

$$\begin{aligned} \mathscr{E}_s &= \text{E}\left\{|y_s[m]|^2\right\} = \text{E}\left\{\left|a[m]\int_{-\infty}^{\infty} q(\tau)g(-\tau)d\tau\right|^2\right\} \\ &= \underbrace{\text{E}\left\{|a[m]|^2\right\}}_{\sigma_a^2}\left|\int_{-\infty}^{\infty} q(\tau)g(-\tau)d\tau\right|^2, \end{aligned}$$

where σ_a^2 denotes average symbol energy. Next, notice that
Equation:

$$y_n[m] = y_n(mT) = \int_{-\infty}^{\infty} q(\tau)\widetilde{w}(mT-\tau)d\tau,$$

so that

Equation:

$$\begin{aligned}\mathscr{E}_n &= \mathrm{E}\left\{\left|y_n[m]\right|^2\right\} = \mathrm{E}\left\{\left|\int_{-\infty}^{\infty} q(\tau)\widetilde{w}(mT-\tau)d\tau\right|^2\right\} \\ &= \mathrm{E}\left\{\int_{-\infty}^{\infty} q(\tau)\widetilde{w}(mT-\tau)d\tau \int_{-\infty}^{\infty} q^*(\tau')\widetilde{w}^*(mT-\tau')d\tau'\right\} \\ &= \int_{-\infty}^{\infty} q(\tau)\int_{-\infty}^{\infty} q^*(\tau')\underbrace{\mathrm{E}\left\{\widetilde{w}(mT-\tau)\widetilde{w}^*(mT-\tau')\right\}}_{N_0\delta(\tau'-\tau)}d\tau' d\tau \\ &= N_0\int_{-\infty}^{\infty}\left|q(\tau)\right|^2 d\tau.\end{aligned}$$

Putting these together, we find

Equation:

$$\mathrm{SNR} = \frac{\sigma_a^2}{N_0}\frac{\left|\int_{-\infty}^{\infty} q(\tau)g(-\tau)d\tau\right|^2}{\int_{-\infty}^{\infty}\left|q(\tau)\right|^2 d\tau}.$$

Cauchy-Schwarz says

Equation:

$$\begin{aligned}\left|\int_{-\infty}^{\infty} b(\tau)c(\tau)d\tau\right|^2 &\leq \int_{-\infty}^{\infty}\left|b(\tau)\right|^2 d\tau \cdot \int_{-\infty}^{\infty}\left|c(\tau)\right|^2 d\tau \\ &\quad \text{with equality iff } b(\tau) = Kc^*(\tau) \text{ for any } K,\end{aligned}$$

which implies

Equation:

$$\begin{aligned}\mathrm{SNR} &\leq \frac{\sigma_a^2}{N_0}\int_{-\infty}^{\infty}\left|g(-\tau)\right|^2 d\tau \\ &\quad \text{with equality iff } q(\tau) = Kg^*(-\tau) \text{ for any } K.\end{aligned}$$

Noting that SNR doesn't depend on K, we choose $K = 1$. Thus, given pulse $g(t)$, the SNR-maximizing receiver filter is

Equation:

$$q(\tau) = g^*(-\tau) \quad \text{known as a "matched filter".}$$

We can write this in the frequency domain as

Equation:

$$\begin{aligned} Q(f) &= \int_{-\infty}^{\infty} q(\tau) e^{-j2\pi f\tau} d\tau = \int_{-\infty}^{\infty} g^*(-\tau) e^{-j2\pi f\tau} d\tau \\ &= \int_{-\infty}^{\infty} g^*(t) e^{j2\pi ft} dt = \left[\int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} dt\right]^* \\ &= G^*(f). \end{aligned}$$

Summary: For SNR-maximizing ISI-free pulses, we need

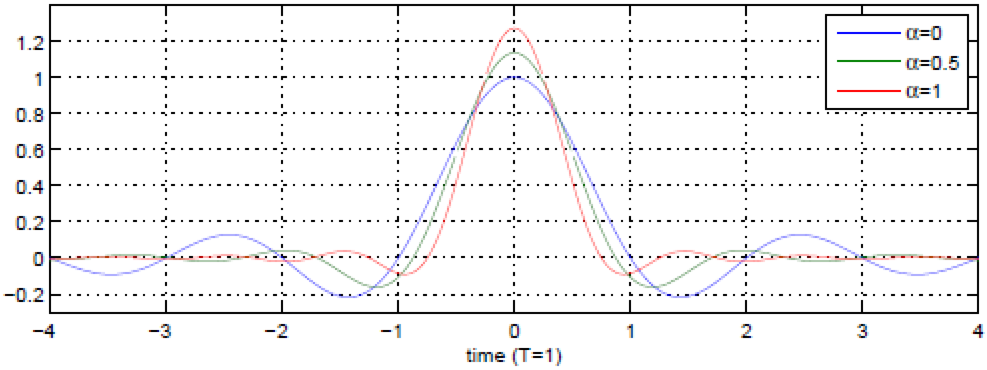
1. $G(f)Q(f) = P(f)$ satisfying the Nyquist criterion,
2. $Q(f) = G^*(f)$,

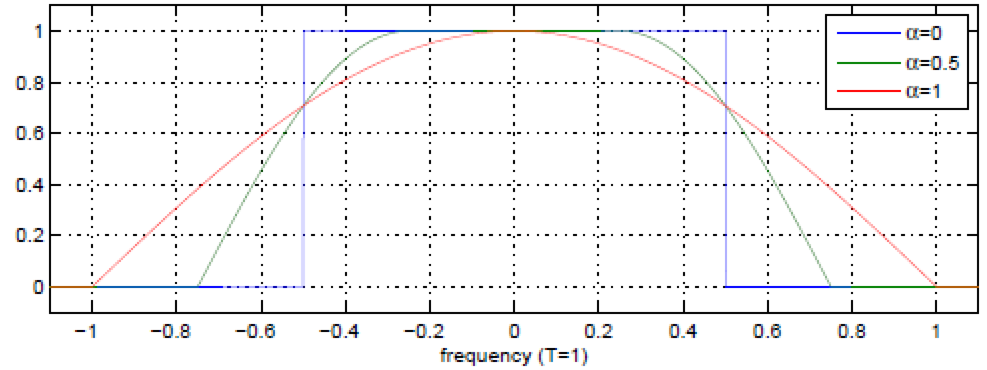
which together imply $|G(f)|^2$ must satisfy the Nyquist criterion.

One option is $G(f) = \sqrt{P_{\text{RC}}(f)}$, since $P_{\text{RC}}(f)$ was Nyquist. We call this the "square-root raised cosine" (SRRC) pulse. Working out the details of $\mathscr{F}^{-1}\{G_{\text{SRRC}}(f)\}$, we find

Equation:

$$g_{\text{SRRC}}(t) = \frac{(1-\alpha)\,\text{sinc}\left(\frac{t}{T}(1-\alpha)\right)}{1-\left(4\alpha\frac{t}{T}\right)^2} + \frac{4\alpha\cos\left(\pi\frac{t}{T}(1+\alpha)\right)}{\pi\left(1-\left(4\alpha\frac{t}{T}\right)^2\right)}.$$



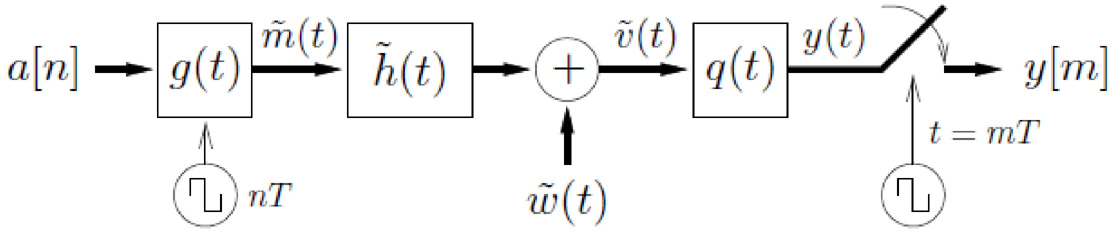


At the receiver, we would use $q_{\text{SRRC}}(t) = g^{*}_{\text{SRRC}}(-t) = g_{\text{SRRC}}(t)$; the latter equality is due to $g_{\text{SRRC}}(t)$ being real and symmetric.

Discrete-Time Implementation of Digital Communication

In this module, we derive the discrete-time implementation of a (baseband-equivalent) digital communication system, which is useful for Matlab simulation and error-rate analysis. To do this, we use the fundamental DSP tool known as "sinc reconstruction", which says that a bandlimited waveform can be represented exactly by a sum of shifted sinc pulses. Sinc reconstruction is separately applied to the transmission pulse, the reception pulse, and the baseband equivalent channel impulse response, after which those three steps can be represented by discrete-time filtering operations. For this, we show that it is essential to sample the filters at least twice as fast as the symbol rate.

Digital implementation of transmitter pulse-shaping and receiver filtering is much more practical than analog. First, recall the analog implementation of our digital communication system: from the module Digital Communication.



To proceed further, we need an important DSP concept called “sinc reconstruction”:

Equation:

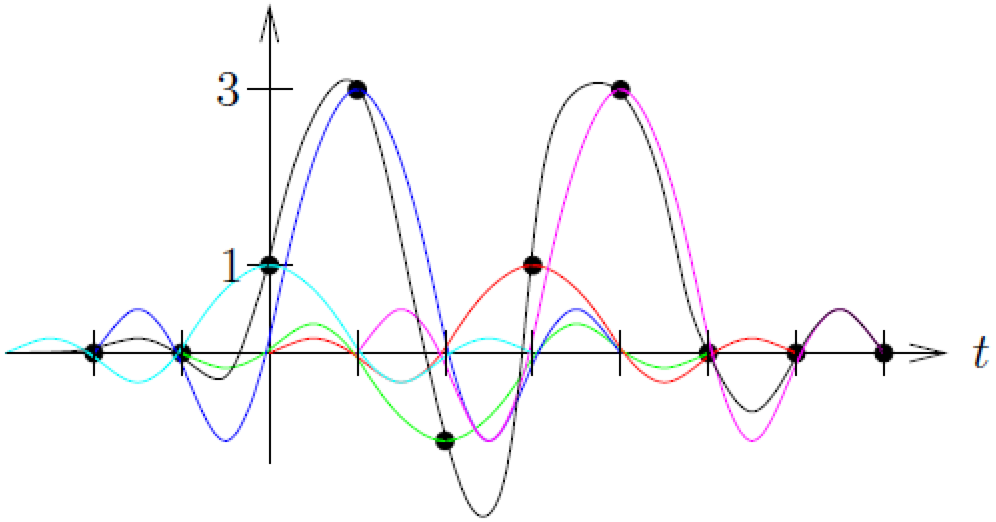
$$\text{If waveform } x(t) \text{ is bandlimited to } \frac{1}{2T_s}\text{Hz, then}$$

Equation:

$$x\left(t\right) = \sum_{n=-\infty}^{\infty} x\left[n\right] \operatorname{sinc}\left(\frac{1}{T_s}(t-nT_s)\right) \text{ for } x\left[n\right] = x\left(nT_s\right).$$

In other words, a bandlimited waveform can be reconstructed from its samples via sinc pulse shaping.

$$x(t) = \sum_n x[n]\,\mathrm{sinc}(\tfrac{1}{T}(t-nT))$$



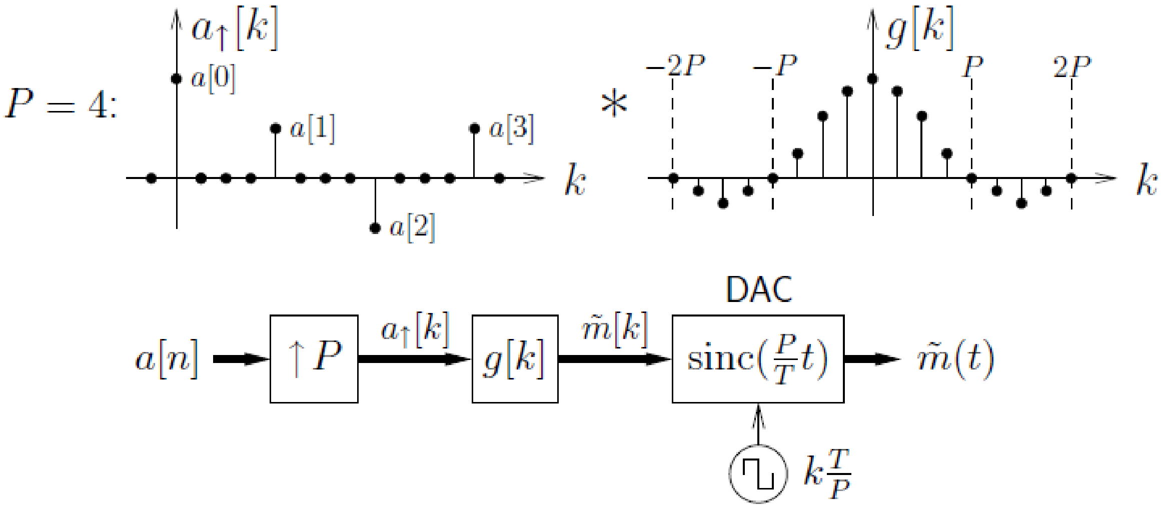
Discrete-Time Pulse-Shaping

Applying $\frac{T}{P}$-sampling and reconstruction to $g(t)$ (where the SRRC pulse bandwidth $\frac{1+\alpha}{2T}$ requires the use of $P \geq 2$),

Equation:

$$\begin{aligned}
g(\tau) &= \textstyle\sum_l g[l]\ \mathrm{sinc}\left(\frac{P}{T}\left(\tau - l\frac{T}{P}\right)\right) \quad \text{for } g[l] = g\left(l\frac{T}{P}\right) \\
\widetilde{m}(t) &= \textstyle\sum_n a[n] g(t-nT) \\
&= \textstyle\sum_n a[n] \sum_l g[l]\ \mathrm{sinc}\left(\frac{P}{T}\left(t - nT - l\frac{T}{P}\right)\right) \\
&= \textstyle\sum_n a[n] \sum_k g[k-nP]\ \mathrm{sinc}\left(\frac{P}{T}\left(t - k\frac{T}{P}\right)\right) \ \text{via } k = nP - l \\
&= \textstyle\sum_k \sum_n a[n] g[k-nP]\ \mathrm{sinc}\left(\frac{P}{T}\left(t - k\frac{T}{P}\right)\right) \\
&\qquad\qquad := \widetilde{m}[k]
\end{aligned}$$

The sequence $\widetilde{m}[k]$, a weighted sum of P-shifted pulses $g[k]$, can be generated by P-upsampling $a[n]$ (i.e., inserting $P-1$ zeros between every pair of samples) and filtering with $g[k]$:



Note:sinc-pulse shaping$=$digital-to-analog conversion (DAC).

Discrete-Time Receiver Filtering

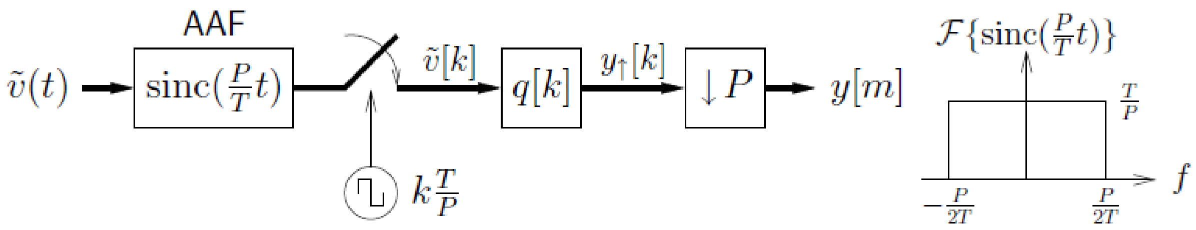
Applying $\frac{T}{P}$-sampling and reconstruction to bandlimited $q(\tau)$:
Equation:

$$q(\tau) \quad = \quad \sum_{l=-\infty}^{\infty} q\,[l]\ \text{sinc}\left(\frac{P}{T}\left(\tau - l\frac{T}{P}\right)\right) \quad \text{for}\ \ q\,[l] = q\left(l\frac{T}{P}\right)$$

where again we need $P \geq 2$, yields
Equation:

$$
\begin{aligned}
y_\uparrow[k] &= y\left(k\tfrac{T}{P}\right) = \int_{-\infty}^{\infty} q(\tau)\,\tilde{v}\left(k\tfrac{T}{P}-\tau\right)d\tau \\
&= \int_{-\infty}^{\infty}\sum_{l=-\infty}^{\infty} q[l]\;\text{sinc}\left(\tfrac{P}{T}\left(\tau - l\tfrac{T}{P}\right)\right)\tilde{v}\left(k\tfrac{T}{P}-\tau\right)d\tau \\
&= \sum_{l=-\infty}^{\infty} q[l]\int_{-\infty}^{\infty}\text{sinc}\left(\tfrac{P}{T}\tau'\right)\tilde{v}\left((k-l)\tfrac{T}{P}-\tau'\right)d\tau' \\
&\qquad\qquad \left\{\text{sinc}\left(\tfrac{P}{T}t\right)*\tilde{v}(t)\right\}_{t=(k-l)\frac{T}{P}} = \tilde{v}[k-l] \\
&= \sum_{l=-\infty}^{\infty} q[l]\tilde{v}[k-l] = q[k]*\tilde{v}[k],
\end{aligned}
$$

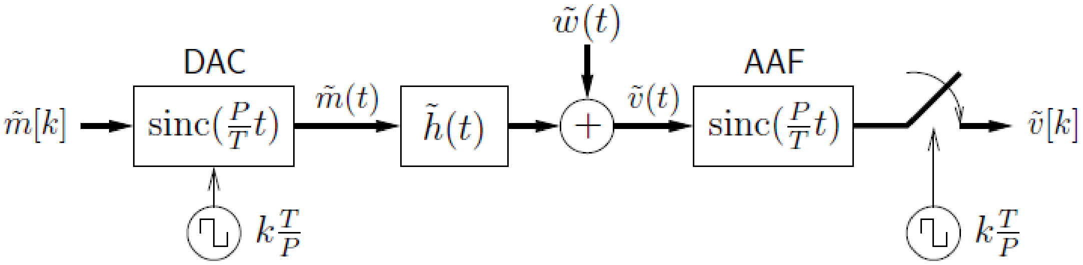
from which $y[m]$ is obtained by keeping only every P^{th} sample:



i.e., downsampling. Here sinc $\left(\frac{P}{T}t\right)$ does anti-alias filtering.

Discrete-Time Complex-Baseband Channel

Finally, we derive a discrete-time representation of the channel between $\widetilde{m}[k]$ and $\tilde{v}[k]$:



Using $\widetilde{w}[k]$ to refer to the noise component of $\tilde{v}[k]$, it can be seen from the block diagram that

Equation:

$$\widetilde{w}[k] = \int_{-\infty}^{\infty} \widetilde{w}(\tau)\,\mathrm{sinc}\left(\frac{P}{T}\left(k\frac{T}{P} - \tau\right)\right)d\tau$$

To model the signal component of $\tilde{v}[k]$, realize that $\widetilde{m}[k]$ is effectively pulse-shaped by $\mathrm{sinc}\left(\frac{P}{T}t\right) * \widetilde{h}(t) * \mathrm{sinc}\left(\frac{P}{T}t\right)$. But since the frequency response of $\mathrm{sinc}\left(\frac{P}{T}t\right)$ has a flat gain of $\frac{T}{P}$ over the signal bandwidth, and thus the bandwidth of $\widetilde{h}(t)$,

Equation:

$$\mathrm{sinc}\left(\frac{P}{T}t\right) * \widetilde{h}(t) * \mathrm{sinc}\left(\frac{P}{T}t\right) = \frac{T^2}{P^2}\widetilde{h}(t).$$

So, with $\frac{T}{P}$-sampling and reconstruction of $\frac{T^2}{P^2}\widetilde{h}(t)$, i.e.,

Equation:

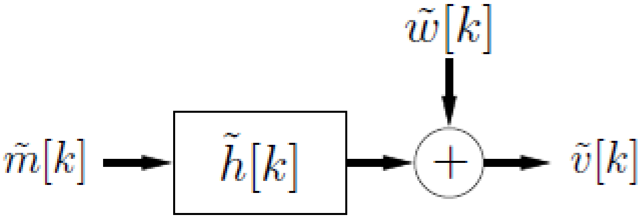
$$\frac{T^2}{P^2}\widetilde{h}(t) = \sum_{i=-\infty}^{\infty} h[i]\,\mathrm{sinc}\left(\frac{P}{T}\left(t - i\frac{T}{P}\right)\right) \quad \text{for } h[i] = \frac{T^2}{P^2} h\left(i\frac{T}{P}\right)$$

we can write $\tilde{v}[k]$ as

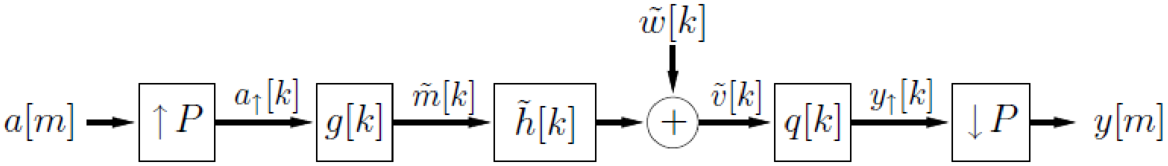
Equation:

$$\begin{aligned}
\tilde{v}[k] &= \tilde{w}[k] + \sum_l \tilde{m}[l] \frac{T^2}{P^2} \tilde{h}\left(k\frac{T}{P} - l\frac{T}{P}\right) \\
&= \tilde{w}[k] + \sum_l \tilde{m}[l] \sum_i \tilde{h}[i]\mathrm{sinc}\,(k-l-i) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \delta[k-l-i] \\
&= \tilde{w}[k] + \sum_l \tilde{m}[l]\tilde{h}[k-l]
\end{aligned}$$

yielding the discrete-time channel



Merging the discrete-time channel with the discrete-time modulator and demodulator yields

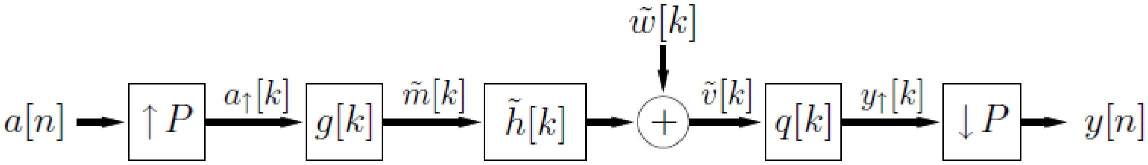


known as the “fractionally sampled” system model. This model is very convenient for MATLAB simulation and acts as a foundation for further analysis.

Error Analysis of Digital Communications

In this module, we first introduce the eye diagram and constellation diagram as qualitative ways of evaluating the symbol error probability of a digital communication system. We discuss various symbol alphabets, such as QAM, PAM, and PSK, and their associated decision regions. Finally, we derive the symbol error probability for PAM and QAM in additive white Gaussian noise, using the Q and erfc functions, and discuss Gray coding.

Recall the figure below, from the module Discrete-Time Implementation of Digital Communication. When the channel is trivial and noiseless and the pulses satisfy the Nyquist criterion (i.e., $g[k]*q[k] = \delta[k]$), the digital comm system will work perfectly, yielding $y[n] = a[n]$.



In practice, however,

- the pulses $g[k]$ and $q[k]$ will be truncated to finite length,
- the channel will not be trivial (i.e., $\tilde{h}\,[k] \neq \delta\,[k]$), and
- the channel will not be noiseless (i.e., $\tilde{w}\,[k] \neq 0$),

leading to $y[n] \neq a[n]$, in which case we must *infer* the value of $a[n]$ from the received samples $\{y\,[m]\}_{m=-\infty}^{\infty}$. For now, we consider using only the single sample $y[n]$ to infer $a[n]$.

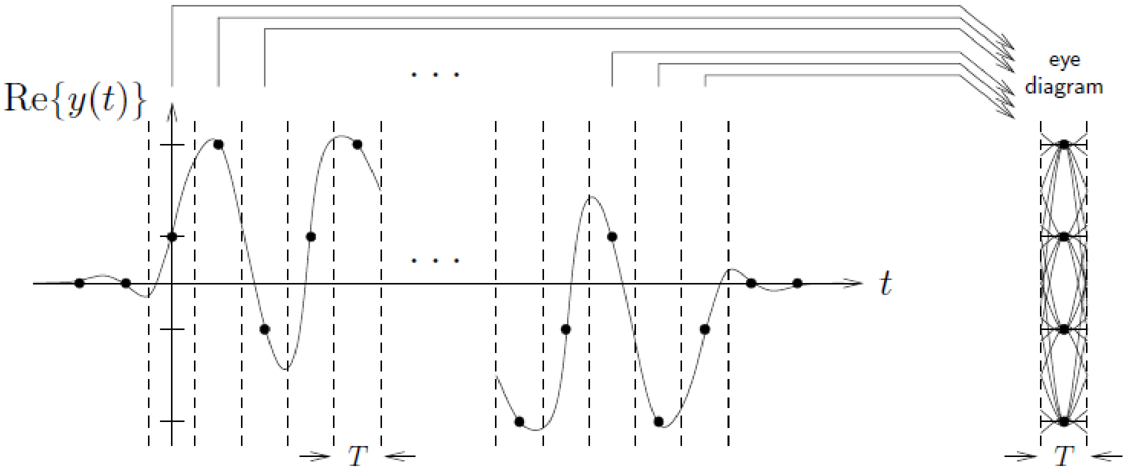
Key question: What are the mechanisms by which errors are made?

To better understand error behavior, we can plot the “eye diagram” or the “constellation diagram” and calculate the symbol error rate (SER).

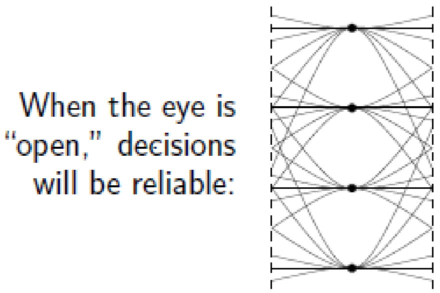
Eye Diagrams

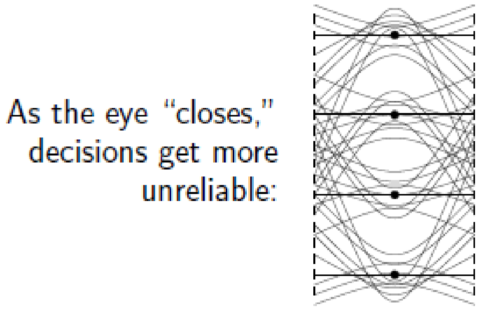
Usually used when $a[n] \in \mathbb{R}$, the eye diagram is a plot which superimposes T-second segments of $\operatorname{Re}\,\{y(t)\}$ over the time intervals

$t \in \left[nT - \frac{T}{2}, nT + \frac{T}{2}\right)$ for many integers n.



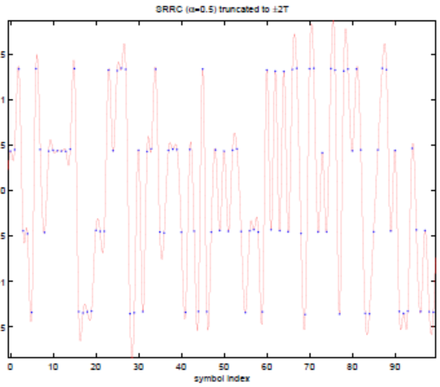
In MATLAB, the eye diagram can be made by superimposing P-sample segments of $\operatorname{Re}\{y_{\uparrow}[k]\}$ corresponding to intervals $k \in \left\{nP - \frac{P}{2}, \cdots, nP + \frac{P-1}{2}\right\}$ for many n. (Usually $P \geq 8$.)

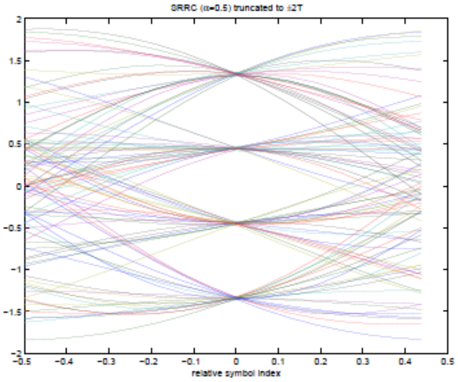




If $a[n] \in \mathbb{C}$, eye diagrams can be plotted for both the "I" and "Q" channels using $\operatorname{Re}\{y(t)\}$ and $\operatorname{Im}\{y(t)\}$, respectively. MATLAB for digital mod/demod with eye diagram:

```
"% design SRRC
P = 16; % oversampling factor
alpha = 0.5; % SRRC rolloff
param
D = 2; % truncation to [-DT,DT]
g = srrc(D,alpha,P); % SRRC
pulse
Ng = length(g);
% generate symbols
N = 100; % # symbols
M = 4; % alphabet size
sig2a = 1; % symbol variance
a = pam(N,M,sig2a); % symbol
sequence
% pulse-amplitude modulate
a_up = zeros(1,N*P);
a_up(1:P:end) = a; % upsampled
symbols
m = conv(a_up,g); % PAM
% matched-filter demodulate
y_up = conv(m,g); % use SRRC
again
% remove causal filtering delay
left = -P/2; % left edge of eye
diagram
k = left+[1:P*N]; % desired
time indices
dly = (Ng-1)/2+(Ng-1)/2;% delay
due to pulses
y_up = y_up(k+dly); % remove
delay
y = y_up([1-left:P:P*N]);%
downsample
% plot received signal
figure(1)
plot(left/P+[0:P*N-
1]/P,y_up,’r’,[0:N-1],y,’.’);
axis(‘tight’)
```

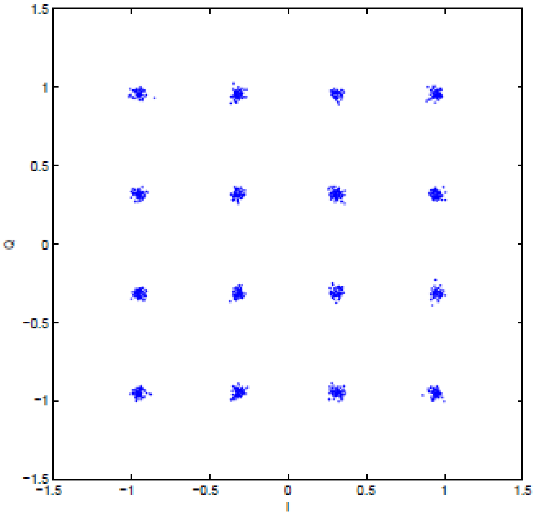


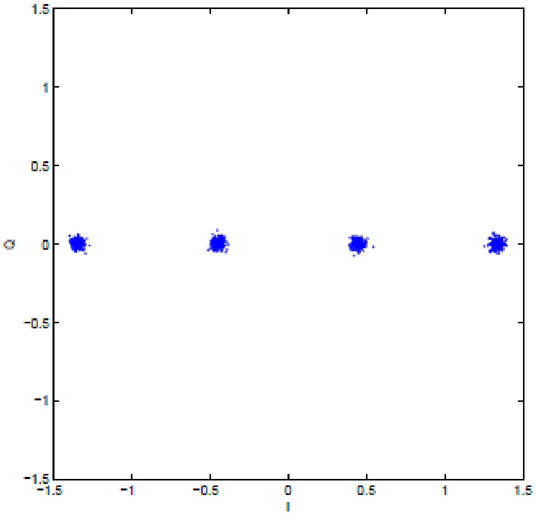


```
title([‘SRRC
(\alpha=‘,num2str(alpha),...
‘) truncated to
\pm’,num2str(D),’T’])
xlabel([‘symbol index’])
% plot eye diagram
figure(2)
Y_up = reshape(y_up,P,N); %
extract segments
plot(left/P+[0:P-1]/P,Y_up) %
superimpose them
title([‘SRRC
(\alpha=‘,num2str(alpha),...
‘) truncated to
\pm’,num2str(D),’T’])
xlabel([‘relative symbol
index’])
```

Constellation Diagrams

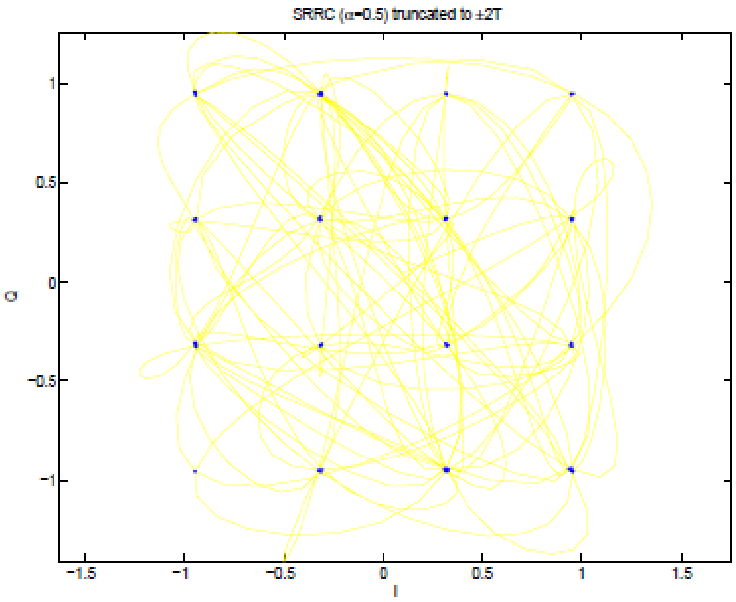
The constellation diagram is a plot of $\mathrm{Im}\,\{y[n]\}$ vs. $\mathrm{Re}\,\{y[n]\}$ for many integers *n*. When the comm system is working well, the points cluster around the symbol alphabet values:





Recall that $y[n] \in \mathbb{C}$ due to the complex-baseband channel model, regardless of whether $a[n] \in \mathbb{R}$ or $a[n] \in \mathbb{C}$.

Sometimes it is instructive to superimpose a plot of $\text{Im}\,\{y_{\uparrow}\,[n]\}$ vs. $\text{Re}\,\{y_{\uparrow}\,[n]\}$, which approximates the trajectory of $y(t)$ in the complex plane:

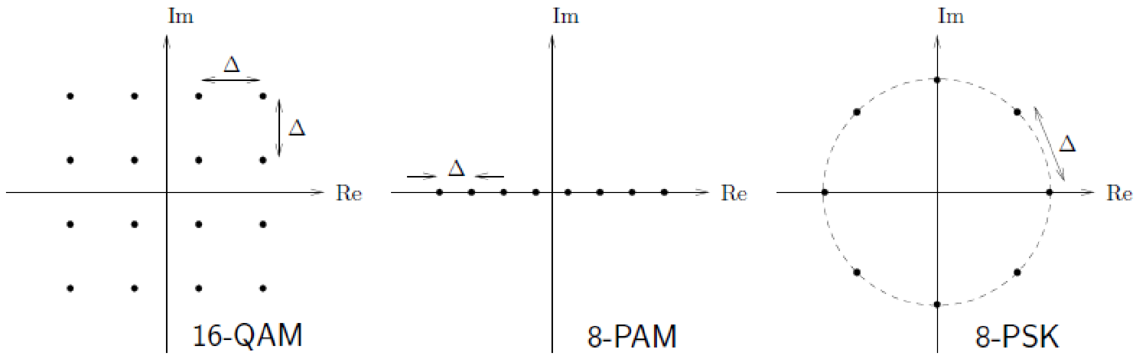


MATLAB for digital mod/demod with constellation diagram:

```
% design SRRC
P = 8; % oversampling factor
alpha = 0.5; % SRRC rolloff param
D = 2; % truncation to [-DT,DT]
g = srrc(D,alpha,P); % SRRC pulse
Ng = length(g);
% generate symbols
N = 100; % # symbols
M = 4; % (sqrt) alphabet size
sig2a = 1; % symbol variance
a = qam(N,M,sig2a); % symbol sequence
% pulse-amplitude modulate
a_up = zeros(1,N*P);
a_up(1:P:end) = a; % upsampled symbols
m = conv(a_up,g); % PAM
% matched-filter demodulate
y_up = conv(m,g); % use SRRC again
% remove causal filtering delay
k = [1:P*N]; % desired time indices
dly = (Ng-1)/2+(Ng-1)/2;% delay due to pulses
y_up = y_up(k+dly); % remove delay
y = y_up(1:P:end); % downsample
% plot received signal
figure(1)
plot3([0:P*N-1]/P,real(y_up),imag(y_up),...
‘r’,[0:N-1],real(y),imag(y),’.’);
xlabel(‘symbol index’);
ylabel(‘I’); zlabel(‘Q’);
view(20,30); % to see full trajectory
%also try view(0,90), view(0,0), view(90,0)
% plot constellation diagram
figure(2)
plot(real(y_up),imag(y_up),’y’,real(y),imag(y),’.’);
xlabel(‘I’); ylabel(‘Q’);
title([‘SRRC (\alpha=‘,num2str(alpha),...
‘) truncated to \pm’,num2str(D),’T’])
axis(‘equal’);
```

Popular Symbol Alphabets

QAM : “quadrature amplitude modulation”
PAM : “pulse amplitude modulation”
PSK : "phase shift keying"



Note that: "QPSK" = 4-QAM = 4-PSK and "BPSK" = 2-PAM = 2-PSK

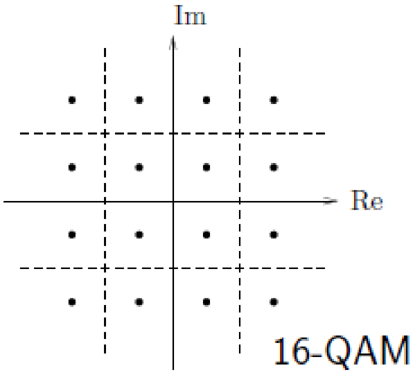
When the alphabet entries are spaced by Δ and picked with equal probability, the symbol variance $\sigma_a^2 = \mathrm{E}\left\{|a[n]|^2\right\}$ obeys:

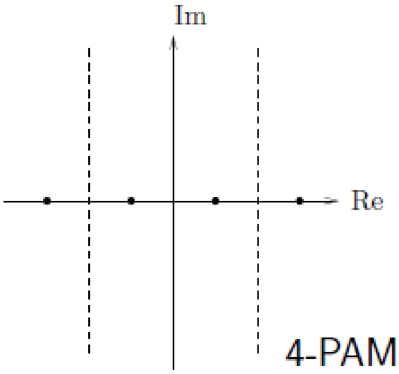
alphabet	M^2-QAM	M-PAM	M-PSK
σ_a^2	$\frac{\Delta^2}{6}\left(M^2 - 1\right)$	$\frac{\Delta^2}{12}\left(M^2 - 1\right)$	$\frac{\Delta^2}{4\sin^2(\pi/M)}$

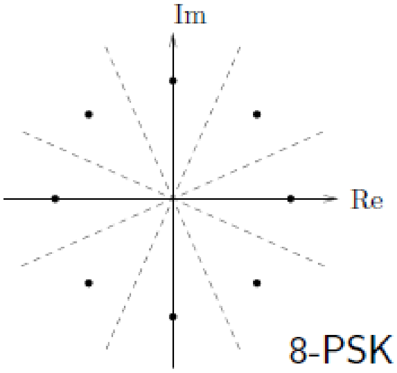
Decision Regions

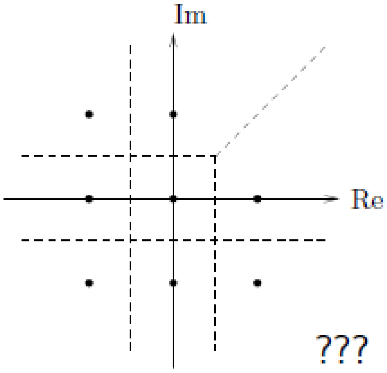
A reasonable way to infer the transmitted symbol $a[n]$ from the received sample $y[n]$ is to decide that $a[n]$ was the alphabet element *nearest* to $y[n]$.

Nearest-element decision making is equivalent to using *decision regions* whose boundaries are equidistant from the two nearest alphabet elements:







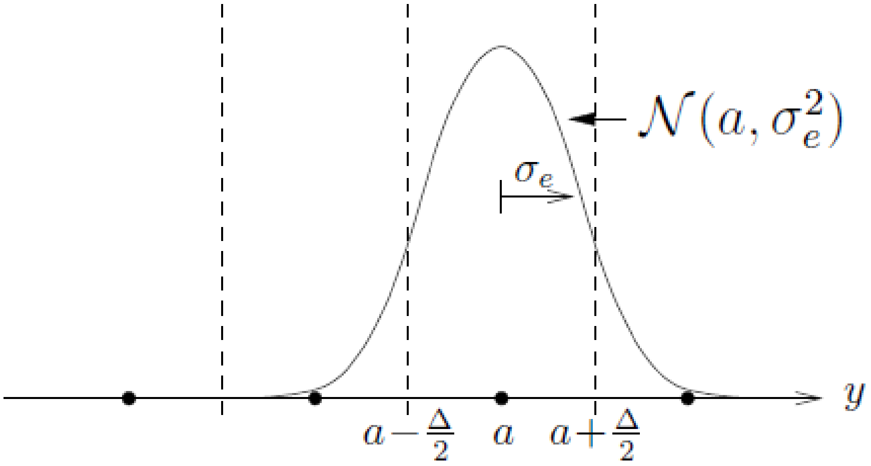


When $a[n] = a$, the *symbol error rate* (SER) equals the probability that $y[n]$ lies outside the decision region corresponding to alphabet member a. Writing $y[n] = a[n] + e[n]$, we represent the cumulative effect of noise and ISI by the error $e[n]$. Usually we model $e[n]$ as a Gaussian random variable with mean 0 and variance σ_e^2.

Symbol Error Rate (SER) for *M*-PAM

Let's first consider an *M*-PAM alphabet, where $a[n] \in \mathbb{R}$. Since the decision regions show that $\operatorname{Im}\{y[n]\}$ is not useful, we'll consider only the real parts of $y[n]$ and $e[n]$.

When $a[n] = a$, we have $y[n] = a + e[n]$, implying that $y[n]$ is Gaussian with mean a and variance $\sigma_e{}^2$, abbreviated as "$\mathcal{N}\left(a, \sigma_e^2\right)$". This is illustrated below for the case of 4-PAM:



Formally, we say that $p_{y[n]|a[n]}(y|a)$, the probability density function (pdf) of $y[n]$ conditioned on $a[n] = a$, obeys

Equation:

$$p_{y[n]|a[n]}(y|a) = \underbrace{\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{(y-a)^2}{2\sigma_e^2}\right)}_{\mathcal{N}(a,\sigma_e^2)}.$$

Basically, $p_{y[n]|a[n]}(y|a)$ *tells us how likely it is that* $y[n] = y$ *given that* $a[n] = a$.

Consider first the case where a is an "interior" (not an "edge") element of the symbol alphabet. Given that $a[n] = a$, we make an error when

$y[n] < a - \frac{\Delta}{2} or$ when $y[n] > a + \frac{\Delta}{2}$. To find the probability of the latter error event, i.e.,

Equation:

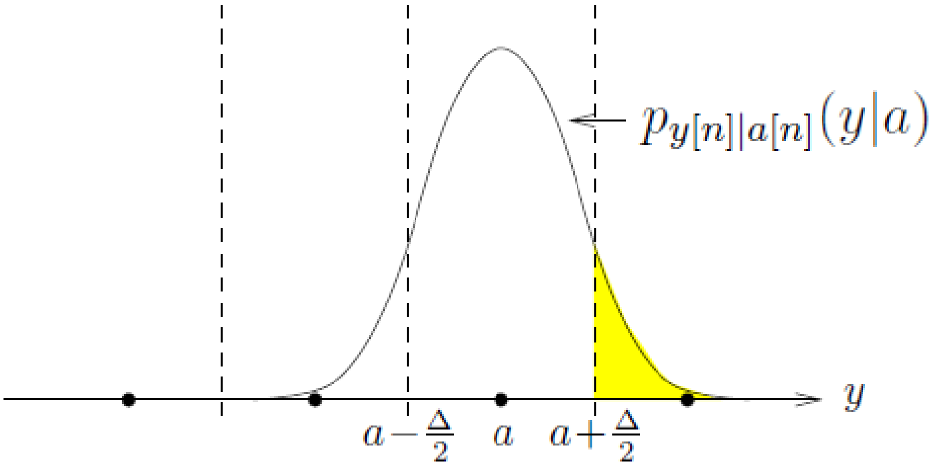
$$\Pr\left\{y[n] > a + \tfrac{\Delta}{2} \mid a[n] = a\right\},$$

we integrate $p_{y[n]|a[n]}(y|a)$ over $y \in \left(a + \frac{\Delta}{2}, \infty\right)$:

Equation:

$$\int_{a+\frac{\Delta}{2}}^{\infty} \underbrace{p_{y[n]|a[n]}(y|a)}_{\mathcal{N}(a,\sigma_e^2)} dy = \int_{a+\frac{\Delta}{2}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{(y-a)^2}{2\sigma_e^2}\right) dy.$$

The integral represents the shaded area below:

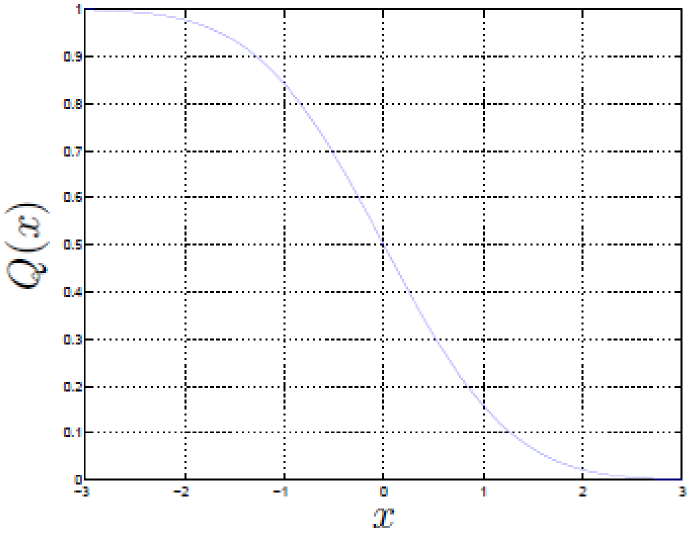


This integral is often solved via

Equation:

$$\int_x^{\infty} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}_{\mathcal{N}(\mu,\sigma^2)} dy = Q\left(\frac{x-\mu}{\sigma}\right),$$

using the “Q function”:



While the Q function is not represented in MATLAB, it can be calculated using the “complementary error function” erfc:
Equation:

$$Q\left(x\right) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right)$$

In any case, the latter error event occurs with probability
Equation:

$$\begin{aligned}\Pr\left\{y[n] > a + \tfrac{\Delta}{2} \mid a[n] = a\right\} &= Q\left(\frac{\left(a+\frac{\Delta}{2}\right) - a}{\sigma_e}\right) \\ &= Q\left(\frac{\Delta}{2\sigma_e}\right).\end{aligned}$$

By symmetry, the former error event probability is also

Equation:

$$\Pr\left\{y[n] < a - \tfrac{\Delta}{2} \mid a[n] = a\right\} = Q\left(\frac{\Delta}{2\sigma_e}\right).$$

Since these two events are disjoint, the probability of making a decision error on an interior symbol equals their sum:

Equation:

$$Q\left(\frac{\Delta}{2\sigma_e}\right) + Q\left(\frac{\Delta}{2\sigma_e}\right) = 2Q\left(\frac{\Delta}{2\sigma_e}\right).$$

For edge symbols, we experience half the decision error probability, since there is only one decision boundary to cross.

Finally, we average over the conditional error probabilities:

Equation:

$$\begin{aligned}\Pr\{\text{error}\} &= \sum_{a \in \text{alphabet}} \Pr\{\text{error} | a[n] = a\} \underbrace{\Pr\{a[n] = a\}}_{=1/M \text{ for all } a} \\ &= Q\left(\frac{\Delta}{2\sigma_e}\right) \cdot \frac{2}{M} + 2Q\left(\frac{\Delta}{2\sigma_e}\right) \cdot \frac{M-2}{M} \\ &= 2\left(\frac{M-1}{M}\right) Q\left(\frac{\Delta}{2\sigma_e}\right),\end{aligned}$$

Using $\sigma_a^2 = \frac{\Delta^2}{12}\left(M^2 - 1\right)$, we can finally write

Equation:

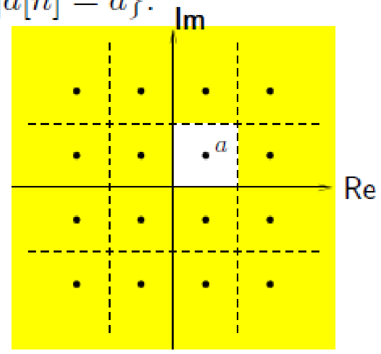
$$\mathrm{SER}_{M\text{-PAM}} = 2\left(\frac{M-1}{M}\right) Q\left(\sqrt{\frac{3}{\left(M^2-1\right)}\frac{\sigma_a^2}{\sigma_e^2}}\right).$$

Symbol Error Rate (SER) for M^2-QAM

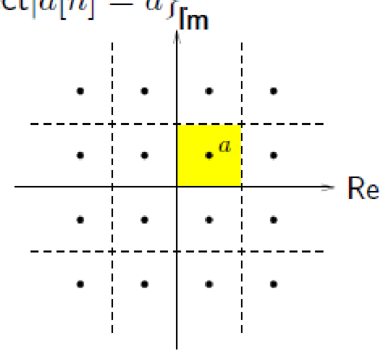
With QAM, we have complex-valued $y[n]$, $a[n]$, $e[n]$. We'll assume that $\mathrm{Re}\,\{e[n]\}$ and $\mathrm{Im}\,\{e[n]\}$ are uncorrelated and equal variance. To calculate SER, we can re-use the PAM approach with a few modifications:

1. integration is done on the complex plane,
2. $\sigma_e{}^2$-variance $e[n] \Rightarrow \frac{\sigma_e^2}{2}$-variance $\mathrm{Re}\,\{e[n]\}$ & $\mathrm{Im}\,\{e[n]\}$,
3. M^2-QAM has 4 corner points, $4(M-2)$ edge points, and $M^2 - 4M + 4$ interior points,
4. calculate $\Pr\,\{\text{error}|a[n] = a\}$ via $1 - \Pr\,\{\text{correct}|a[n] = a\}$, since the regions of integration are simpler:

$\Pr\{\text{error}|a[n] = a\}$:



$\Pr\{\text{correct}|a[n] = a\}$:



After a bit of algebra, we find

Equation:

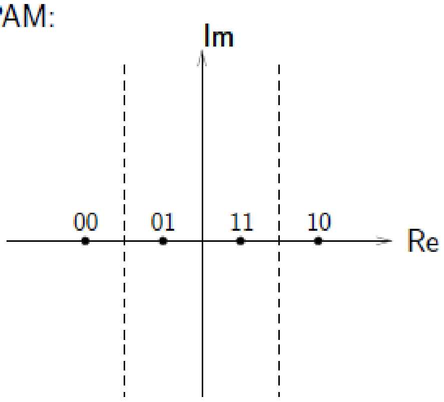
$$\mathrm{SER}_{M^2\text{-QAM}} = 1 - \left[1 - 2\left(\frac{M-1}{M}\right) Q\left(\sqrt{\frac{3}{\left(M^2-1\right)}\frac{\sigma_a^2}{\sigma_e^2}}\right)\right]^2.$$

Bit Error Rate (BER) and Gray Coding

With an M-ary alphabet, there are $\log_2 M$ bits per symbol, so 1 symbol error could cause up to $\log_2 M$ bit errors.

Gray coding is a clever way of mapping bits to symbols so that neighboring symbols differ by only a single bit. Since the vast majority of errors occur when $y[n]$ falls into a neighboring decision region, Gray coding yields BER $\approx$ SER.

4-PAM:



16-QAM: